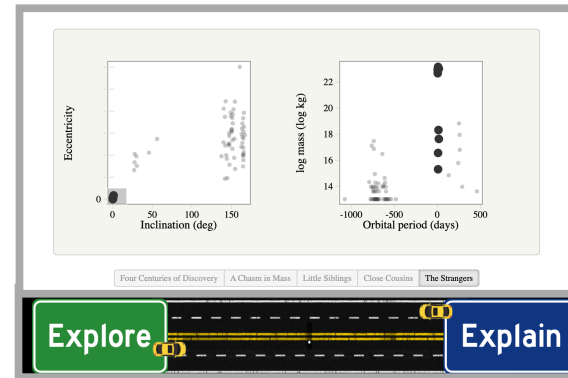
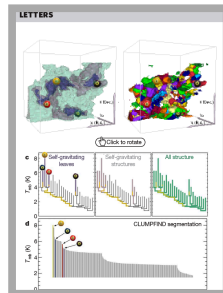
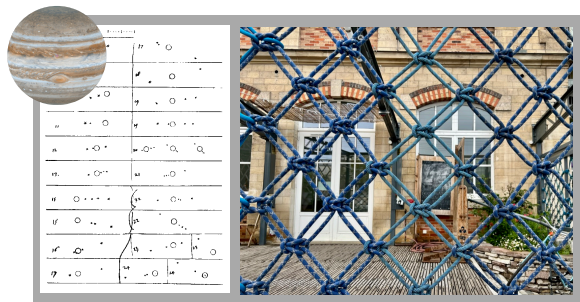


Seeing the universe, more clearly, with **glupyter**

Alyssa A. Goodman • Center for Astrophysics | Harvard & Smithsonian • glue solutions, inc.
with SO many others working on glue, glupyter, and adjacent efforts!

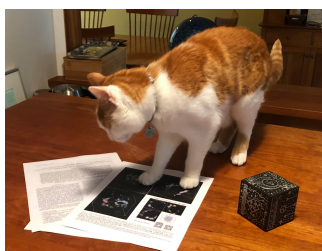
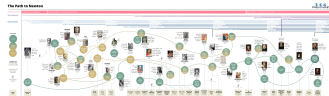
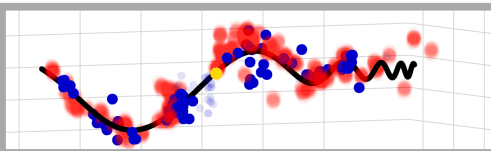




glue

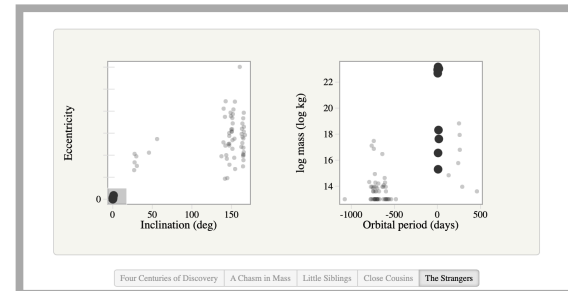
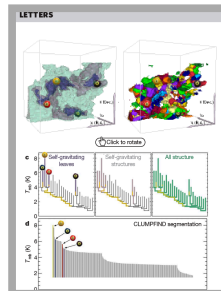
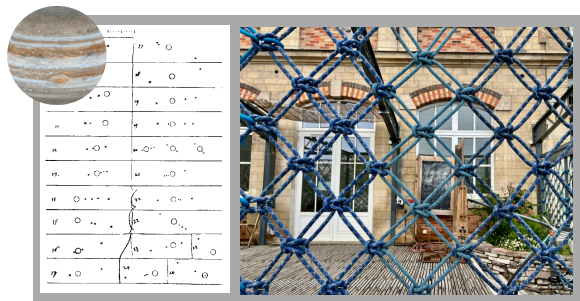
glue

Seeing the universe, more clearly, with **glupyter**

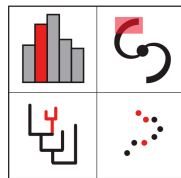



The **TIMELINE** CONSORTIUM





CO SPACES



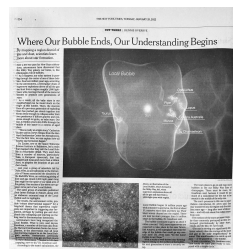
glue
multidimensional data exploration

plotly



glupyter
multidimensional data exploration

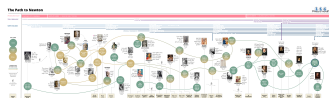
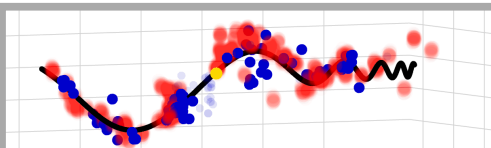
glue genes



CosmicDS



Jdaviz



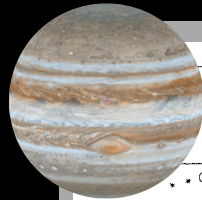
The **TIMELINE**
CONSORTIUM

astrophysics
data system

OpenAI

1 ViZ

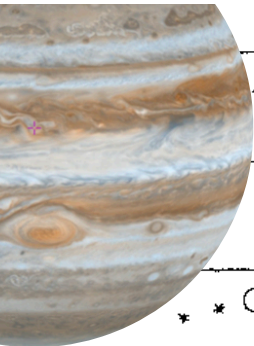
A little bit about astronomy.



A page from Galileo's Sidereus Nuncius showing a table of star positions. The table has two columns of numbers and symbols. The numbers are 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100. The symbols are small circles, some with dots inside, and some with lines through them.



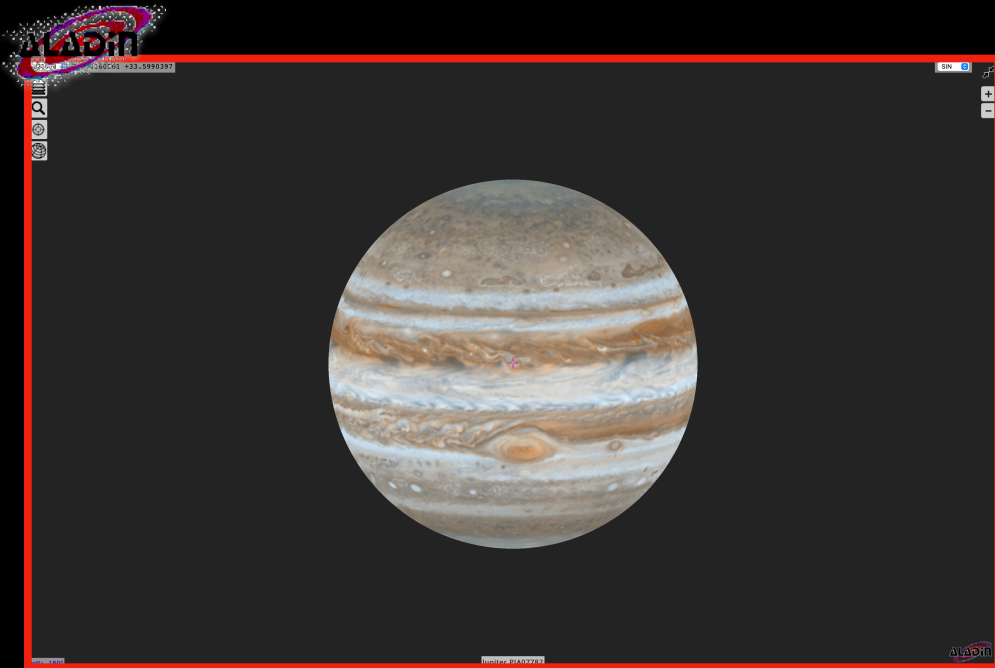




17		17	
*	○	*	○
		*	○
* * ○		* ○ * *	
" * * ○		* ○ * *	
12 * ○ *		20 ○ ○ ○ ○ ○	
13 * ○ ○ ○		21 ... ○ *	
15 ○ * * * *	} 22	* ○ .. *	
18 ○ * * *		22	○ .. * *
16 * ○ *		23 * ○ *	23 ○
17 * ○ *	24	.. ○	24 * ○



Jupiter

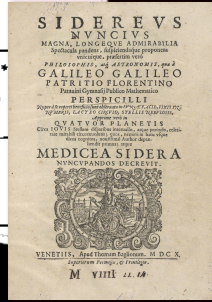
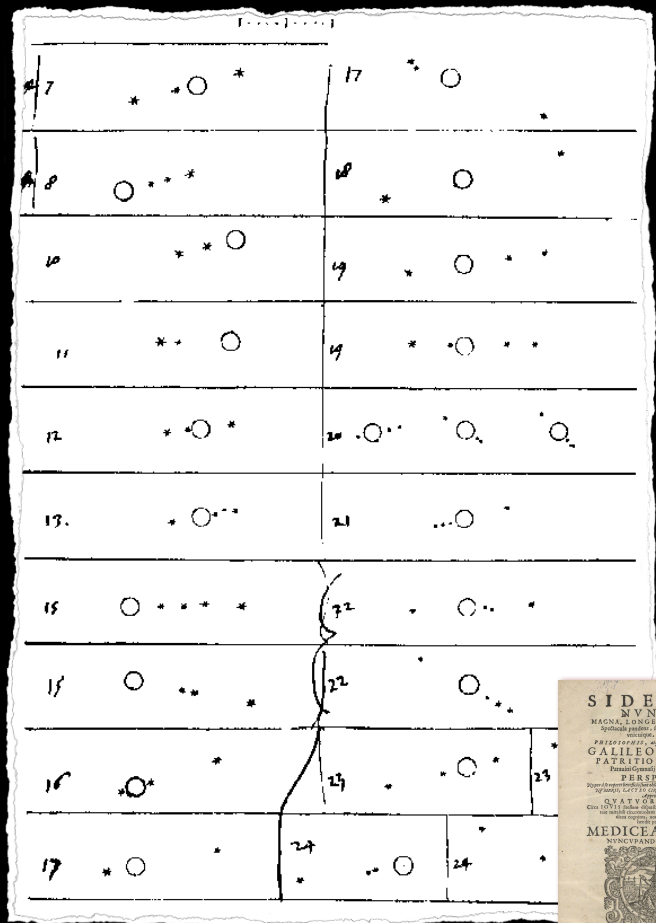


aladin.cds.unistra.fr/AladinLite/planets-explorer/



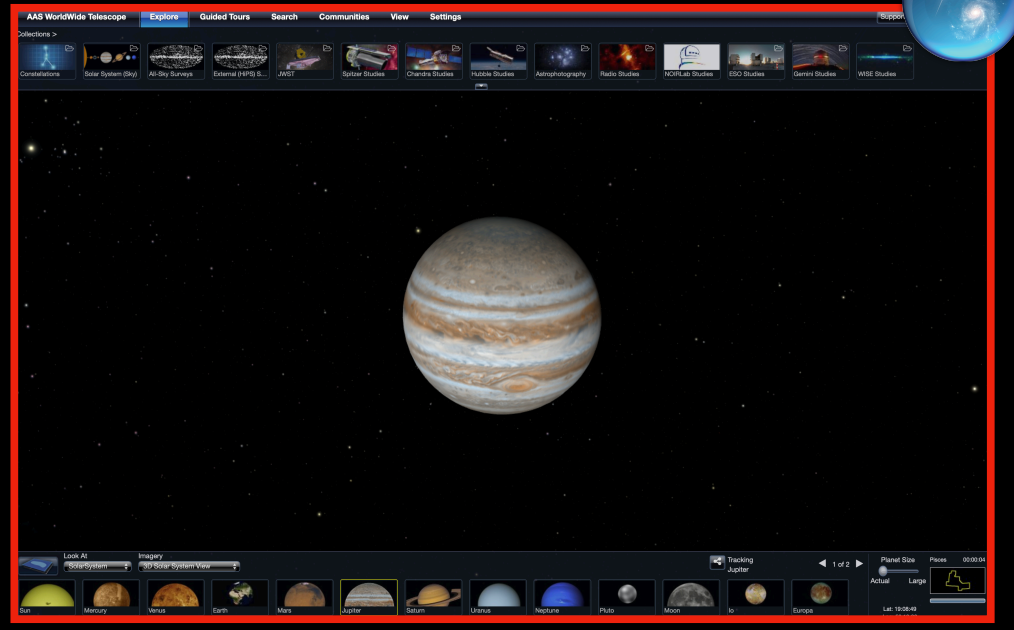
worldwidetelescope.org/webclient/

Galileo, 1610, 2D↔3D caused a revolution



Galileo's "notebook" for Sidereus Nuncius

WWT, today: 2D↔3D easy

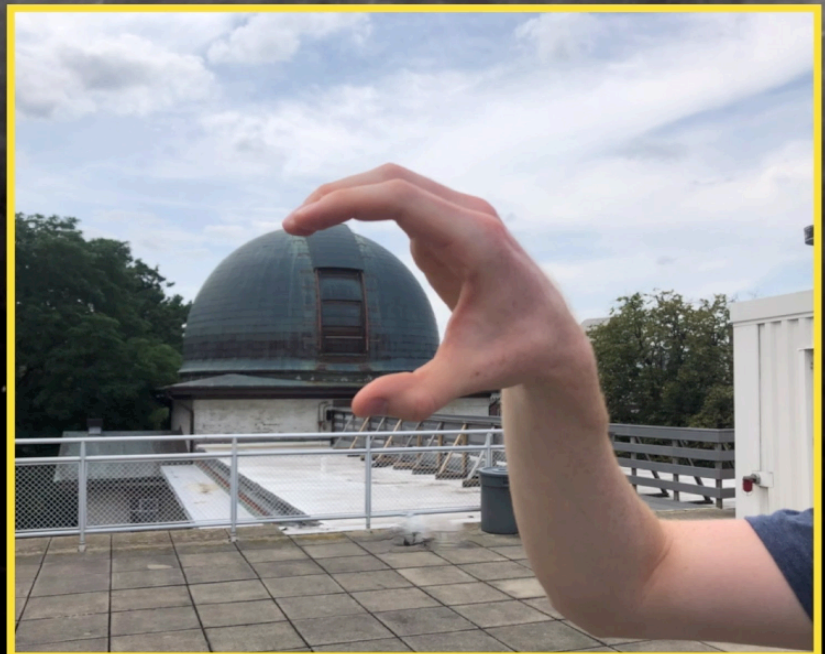


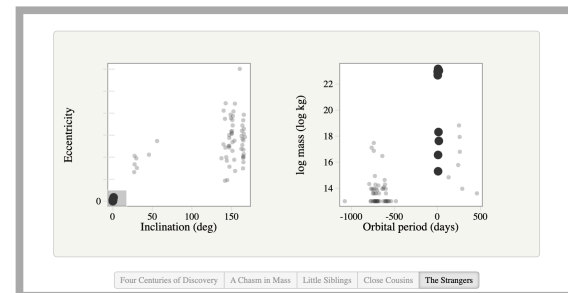
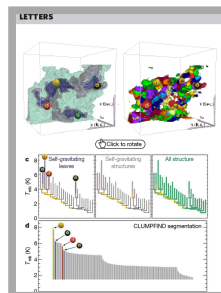
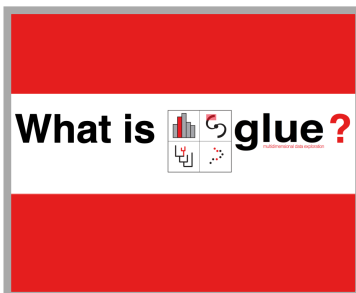
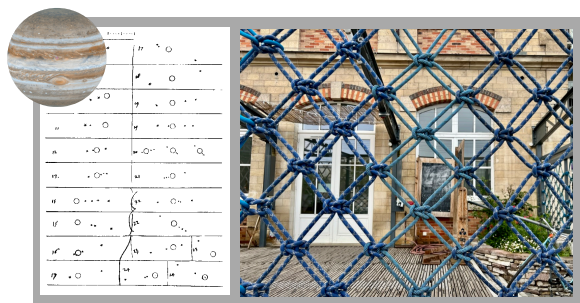
worldwidetelescope.org/webclient/

WWT, today: 2D↔3D easy

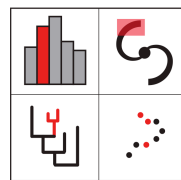
TAURUS

PERSEUS



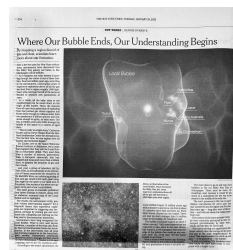


CO SPACES

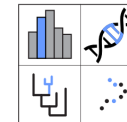


glue
multidimensional data exploration

plotly



glupyter
multidimensional data exploration

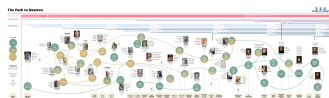
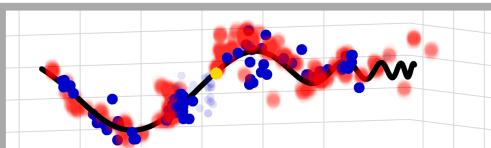


glue
genes

CosmicDS



Jdaviz



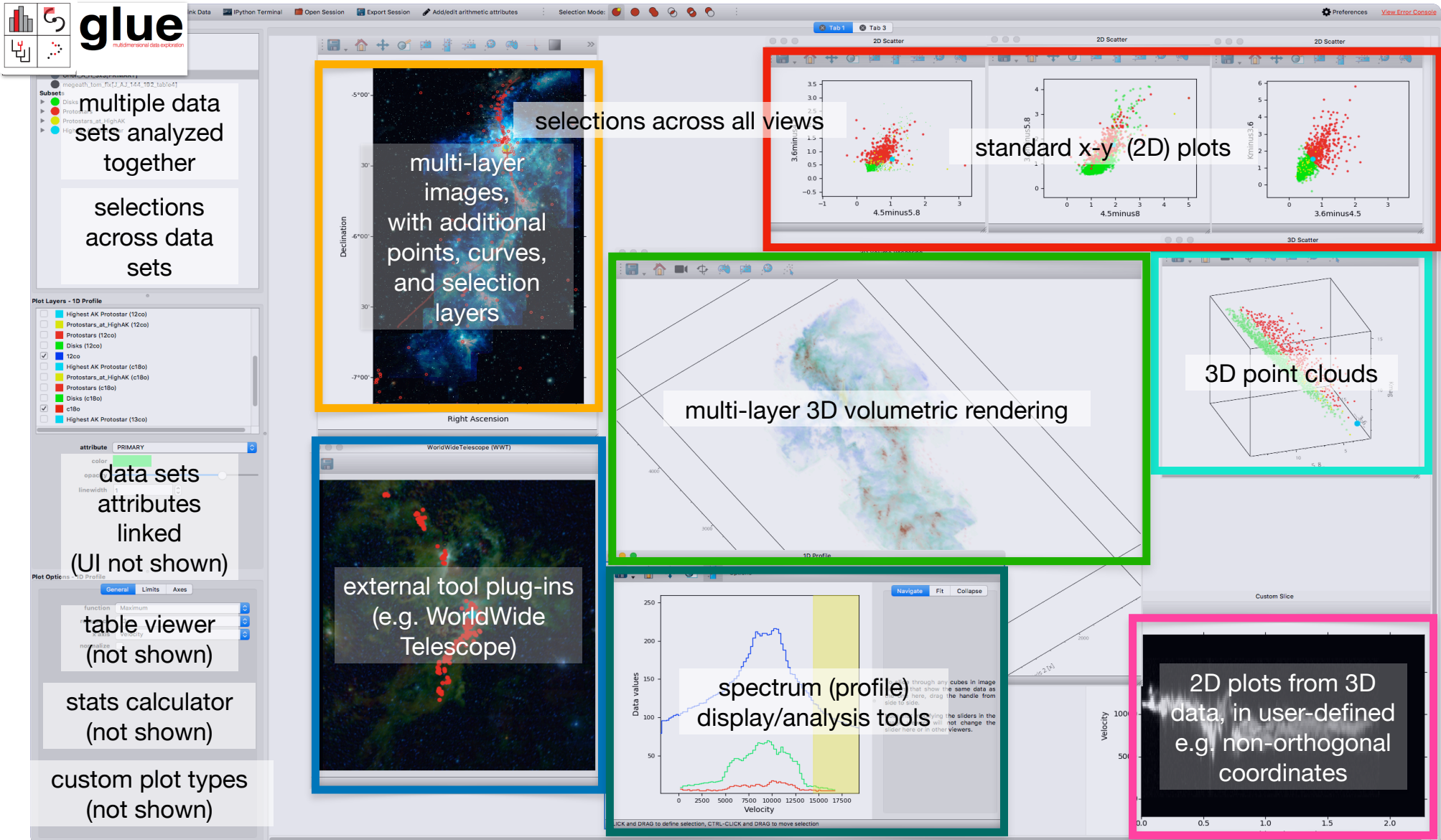
The TIMELINE
CONSORTIUM

astrophysics
data system

OpenAI

1
viz

What is glue?



glue
multidimensional data exploration

multiple data sets analyzed together

selections across data sets

selections across all views

multi-layer images, with additional points, curves, and selection layers

standard x-y (2D) plots

multi-layer 3D volumetric rendering

3D point clouds

data sets attributes linked (UI not shown)

table viewer (not shown)

stats calculator (not shown)

custom plot types (not shown)

external tool plug-ins (e.g. WorldWide Telescope)

spectrum (profile) display/analysis tools

2D plots from 3D data, in user-defined e.g. non-orthogonal coordinates

What is glue?

multidimensional data exploration

It's not an acronym.

It is open-source software that
glues data,
glues graphs &
glues tools.

data



numbers (tables, arrays, spreadsheets)



images & maps (FITS, JPEG, GIS and more)



data cubes (3D, 4D, and more)

data files' common attributes are **glued**



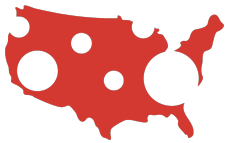
avoiding the need to merge data files

“graphs”



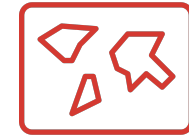
common statistical graphics

(scatterplots, histograms, tables, curves, overlays)



maps & images

(greyscale, color, contours, layer control...)



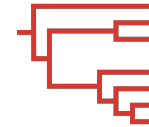
3D displays

(scatter plots, volumetric rendering, sliders...)



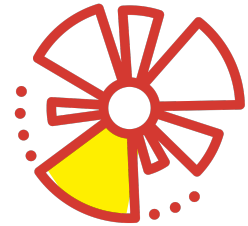
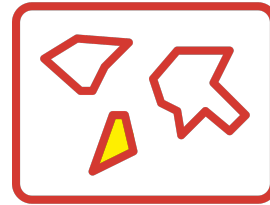
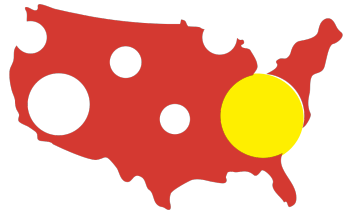
specialized & custom charts

(dendrograms, polar plots, + domain-specific options)





selections propagate across all **graphs**

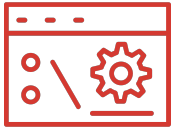


for real-time data exploration & insight

tools



plug-ins (user-defined formats, plots, layouts...)



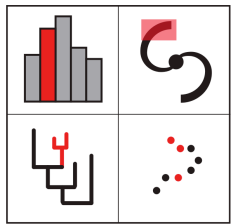
web services (across domains)



command-line (built-in terminal, scriptable)



for easy customization



glue
multidimensional data exploration

glues data,
glues graphs &
glues tools.

glueviz.org

BONUS: **save, share, or publish** what you learn—

save “sessions” to continue where you left off

export graphics

use/export to Jupyter environments

export to plot.ly (javascript)

export to augmented reality

learn how at glueviz.org.



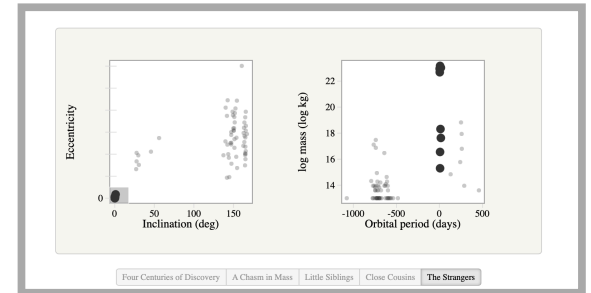
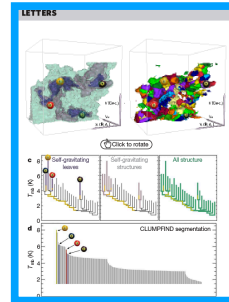
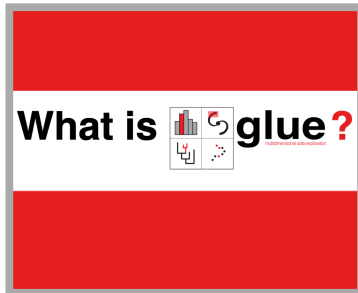
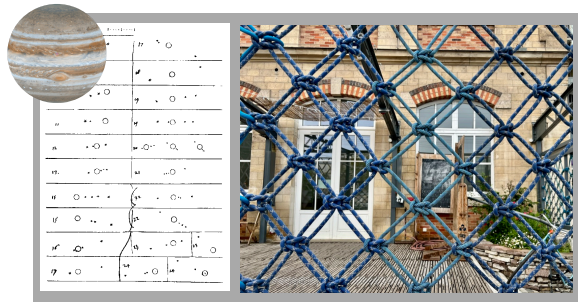
glueviz.org

supported by

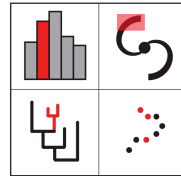


glue
solutions
inc.

GORDON AND BETTY
MOORE
FOUNDATION

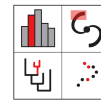


CO SPACES

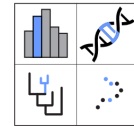


glue
multidimensional data exploration

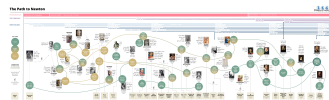
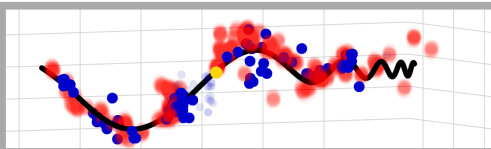
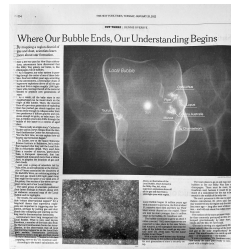
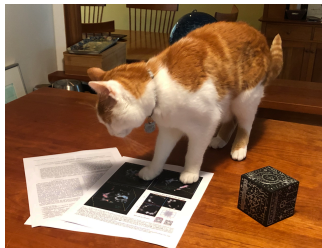
 **plotly**



glupyter
multidimensional data exploration



glue
genes



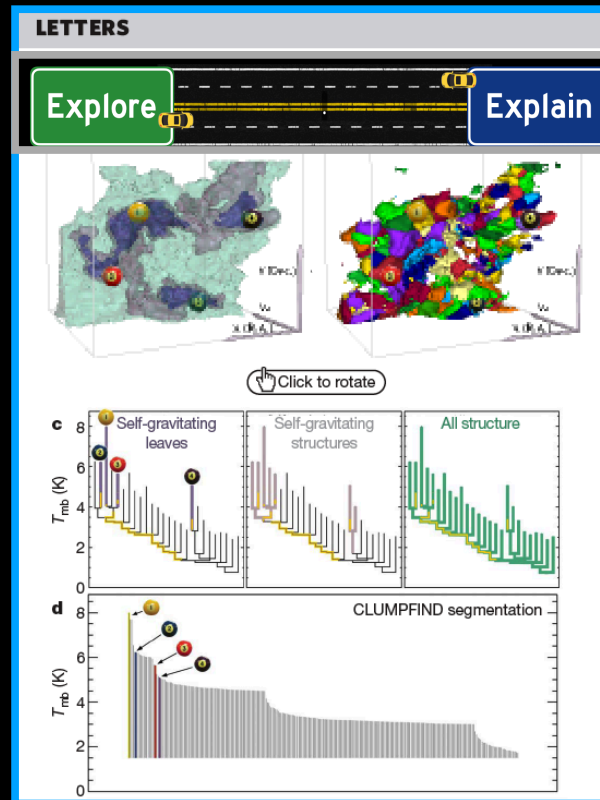
The **TIMELINE**
CONSORTIUM



Origins of glue

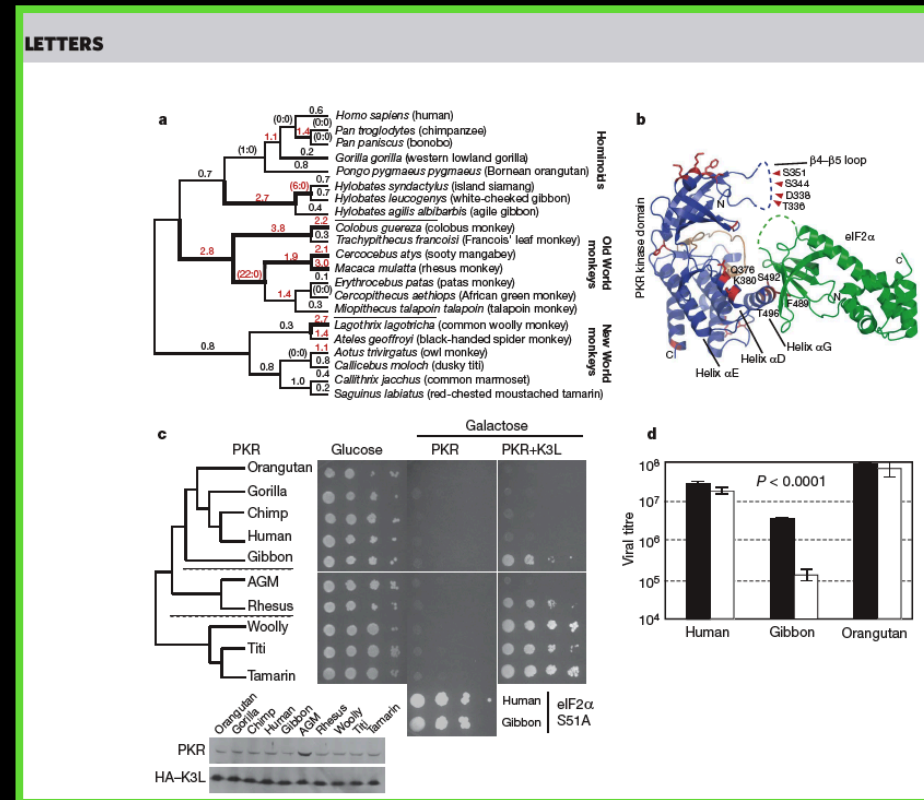
“High-dimensional” or “multivariate” data challenges are similar across astro & bio.

+ we wanted all these plots to talk to each other & be interactive in publications!



Goodman et al. *Nature*, 2009 (IIC)

AstronomicalMedicine@IIC



Elde et al. *Nature*, 2008

Origins of **glupyter**

2005-8

Astronomical**Medicine@iig**

Microsoft Research
WorldWide Telescope

www.worldwidetelescope.org

2008

2011



STScI | SPACE TELESCOPE
SCIENCE INSTITUTE



\$ 2013, launch 2022

Astron. Nachr. / AN 333, No. 5/6, 505 – 514 (2012) / DOI 10.1002/asna.201211705

Principles of high-dimensional data visualization in astronomy

A.A. Goodman*

Harvard-Smithsonian Center for Astrophysics, Cambridge, MA, USA

Received 2012 May 3, accepted 2012 May 4
Published online 2012 Jun 15

Key words cosmology: large-scale structure – ISM: clouds – methods: data analysis – techniques: image processing – techniques: radial velocities

Astronomical researchers often think of analysis and visualization as separate tasks. In the case of high-dimensional data sets, though, interactive *exploratory data visualization* can give far more insight than an approach where data processing and statistical analysis are followed, rather than accompanied, by visualization. This paper attempts to chart a course toward “linked view” systems, where multiple views of high-dimensional data sets update live as a researcher selects, highlights, or otherwise manipulates, one of several open views. For example, imagine a researcher looking at a 3D volume visualization of simulated or observed data, and simultaneously viewing statistical displays of the data set’s properties (such as an x - y plot of temperature vs. velocity, or a histogram of vorticities). Then, imagine that when the researcher selects an interesting group of points in any one of these displays, that the same points become a highlighted subset in all other open displays. Selections can be graphical or algorithmic, and they can be combined, and saved. For tabular (ASCII) data, this kind of analysis has long been possible, even though it has been under-used in astronomy. The bigger issue for astronomy and other “high-dimensional” fields, though, is that no extant system allows for full integration of images and data cubes within a linked-view environment. The paper concludes its history and analysis of the present situation with suggestions that look toward cooperatively-developed open-source modular software as a way to create an evolving, flexible, high-dimensional, linked-view visualization environment useful in astrophysical research.

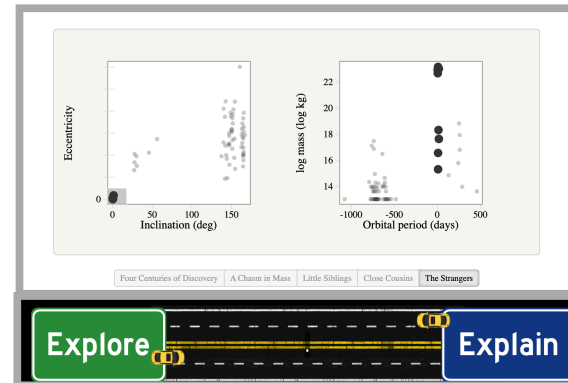
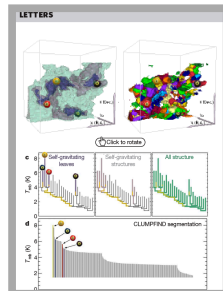
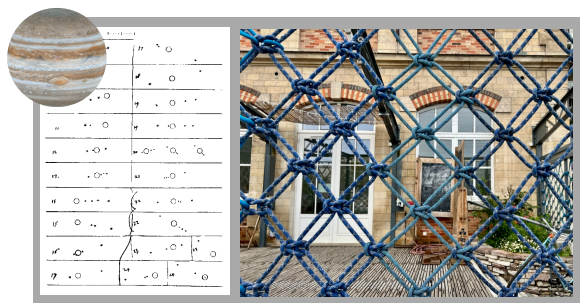
© 2012 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

2012

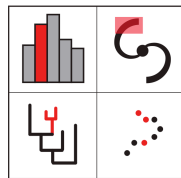
c. 2011



c. 2019



CO SPACES



glue
multidimensional data exploration

plotly



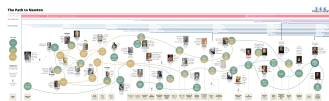
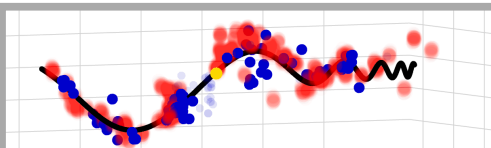
glupyter
multidimensional data exploration

glue
genes

CosmicDS



Jdaviz

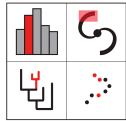


The TIMELINE
CONSORTIUM

astrophysics
data system

OpenAI

1
viz



glue
multidimensional data exploration

c. 2017



enabled by d3.js (javascript) outputs, in 2013

d3po

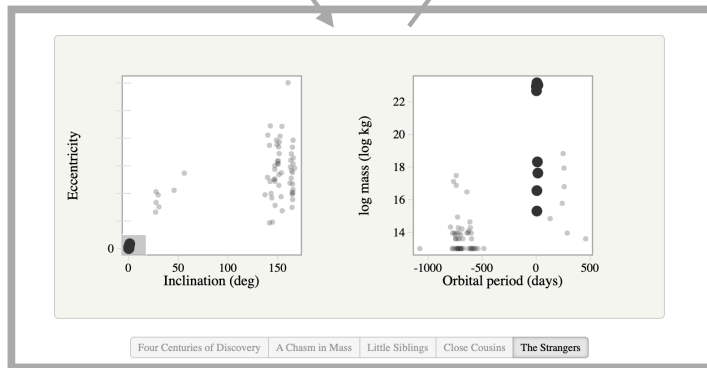
d3po is a project designed to allow an astronomer (or anyone), with no special data visualization skills, to make an interactive, publication-quality figure that has staged builds and linked browsing through scatter plots. Our current version can be previewed at d3po.org, and represents a figure from upcoming work by graduate student Elisabeth Newton. The figure describes how metallicity affects color in cool stars, and represents a nice use case for d3po. Try clicking and dragging in the scatter plots to understand the power of linked browsing in published figures.

Right now we are in search of alpha testers, who have figures that could be made interactive and who are willing to get their hands a little dirty (no javascript skills needed). In future versions, we plan to link to glue to allow the creation of d3po figures interactively. We are also exploring [implementation](#) of d3po within presentations and within [authorea](#). Full 1.0 version expected in January 2014.

Installing your own d3po server

```
git clone git@github.com:adrian/d3po.git
cd d3po
virtualenv --no-site-packages venv
source venv/bin/activate
pip install -r pip-requirements.txt
python run.py
```

Authorea



The "Paper" of the Future
Authorea preprint | 02/21/2017 | DOI: 10.22541/au.148769949.92783646

- Alyssa Goodman (Harvard University)
- Josh Peek (Space Telescope Science Institute)
- Alberto Accomazzi (Harvard-Smithsonian Center for Astrophysics (CFA))
- Chris Beaumont (Harvard-Smithsonian Center for Astrophysics (CFA))
- Christine L. Borgman (UCLA - University of California, Los Angeles)
- Hope-Hwe Hsuan Chen (Harvard University)
- Merce Crosa (Harvard University)
- Christopher Erdmann (North Carolina State University)

And 3 more...

5.5 minute video demonstration of this paper is available at this [YouTube link](#).

1 Preamble
A variety of research on human cognition demonstrates that humans learn and communicate best when more than one processing system (e.g. visual, auditory, touch) is used. And, related research also shows that, no matter how technical the material, most humans also retain and process information best when they can put a narrative "story" to it. So, when considering the future of scholarly communication, we should be careful not to do *bitingly* away with the linear narrative format that articles and books have followed for centuries; instead, we should enrich it.

Much more than text is used to communicate in Science. Figures, which include images, diagrams, graphs, charts, and more, have enriched scholarly articles since the time of Galileo, and ever-growing volumes of data underpin most scientific papers. When scientists communicate face-to-face, as in talks or small discussions, these figures are often the focus of the conversation. In the best discussions, scientists have the ability to manipulate the figures, and to access underlying data, in real-time, so as to test out various what-if scenarios, and to explain findings more clearly. **This short article explains—and shows with demonstrations—how scholarly "papers" can morph into long-lasting rich records of scientific discourse, enriched with deep data and code linkages, interactive figures, audio, video, and commenting.**

Fig. 1
The Paper of the Future should include seamless linkages amongst data, pictures, and language, where "language" includes both words and math. When an individual attempts to understand each of these kinds of information, different cognitive functions are utilized: communication is inefficient if the channel is restricted primarily to language, without easy interconnection to data and pictures.

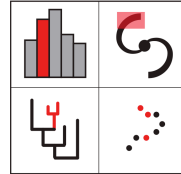
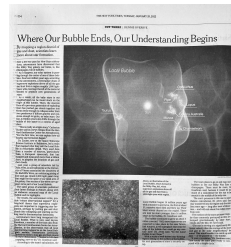
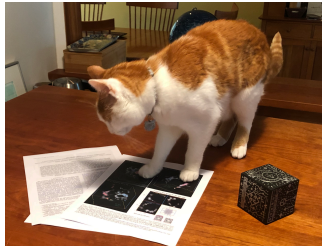
[demo]

Many thanks to Alberto Pepe, Josh Peek, Chris Beaumont, Tom Robitaille, Adrian Price-Whelan, Elizabeth Newton, Michelle Borkin & Matteo Cantiello for making this possible.

Triumphs of **glue** (and plot.ly)

Triumphs of glue (and plot.ly) ++

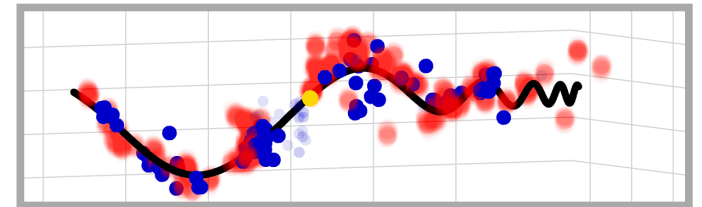
CO SPACES



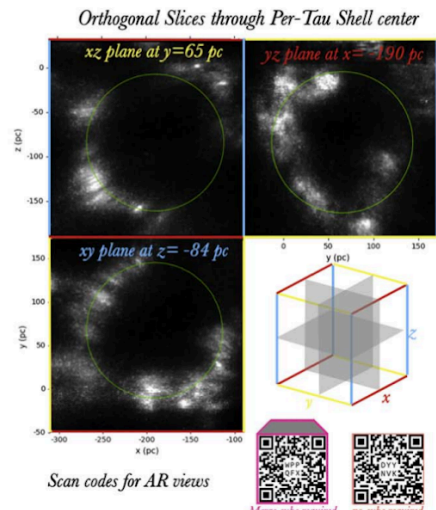
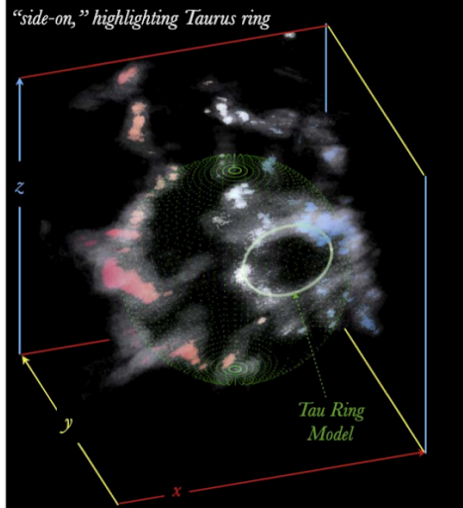
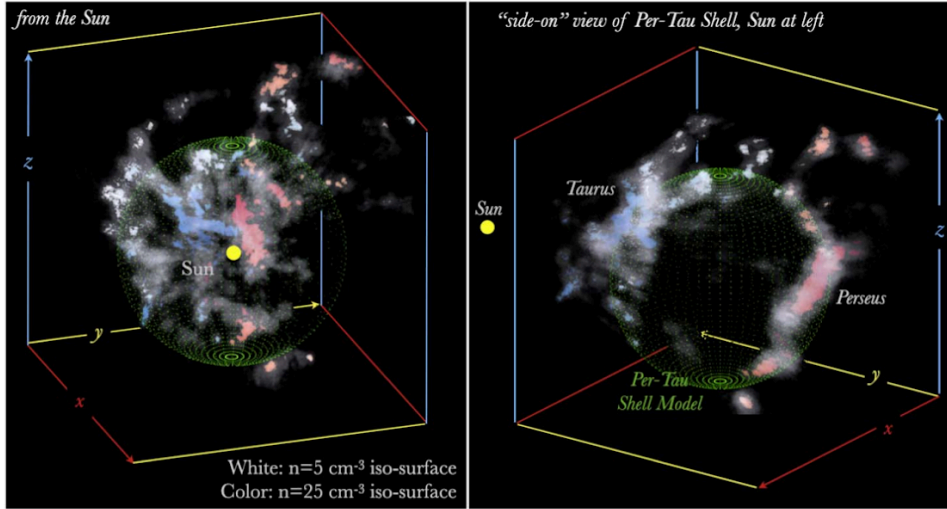
glue
multidimensional data exploration

plotly

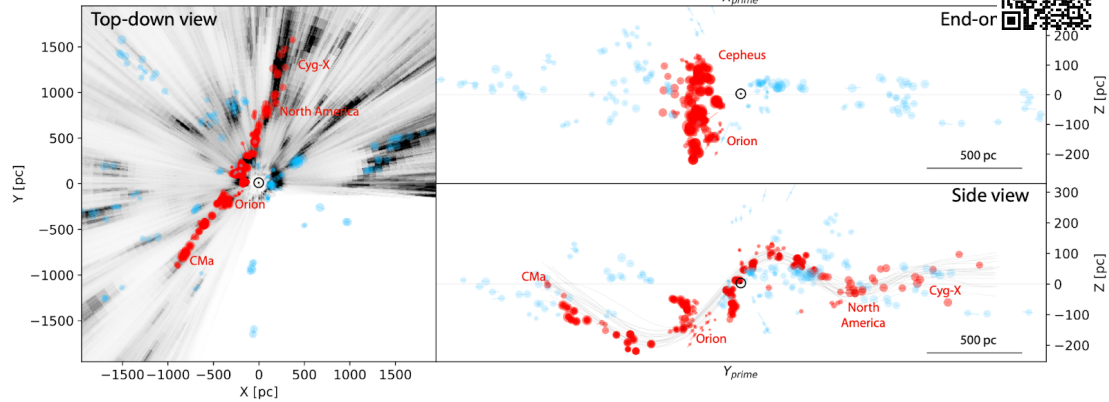
This one is EMBARGOED—I will show you a demo, but please do not post it, or share it.



The Perseus-Taurus Supershell
Bialy et al. 2021, *ApJL*



The Radcliffe Wave, Alves et al. 2020, *Nature*



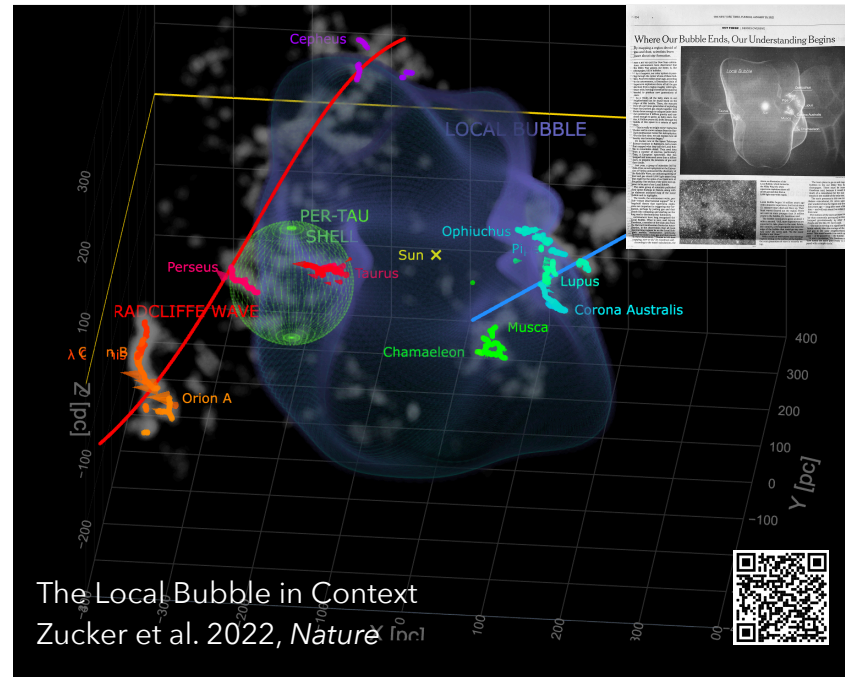
Scan this code for



a “handout”



←AR Codes



Data Collection

Tab 1

3D Scatter

3D Volume Rendering

qt glue: The Radcliffe Wave & the Local Bubble

Data

- Possible Wave Models

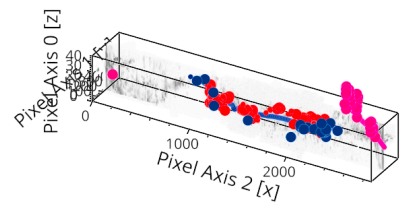
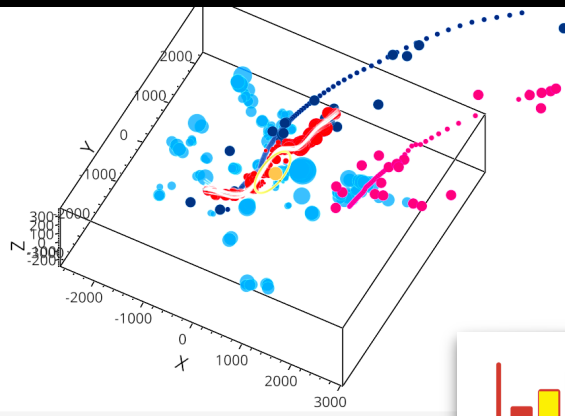
Subsets

- Local Bubble (Reid+2016)
- Major Cloud Catalog
- Maser Catalog (Reid+2014,2016)
- Sagittarius
- Tenuis
- Sun

Plot Layers - 3D

- Radcliff
- Local
- Saggi
- Sun
- Radcliff
- Tenuis

Tenuis Connections



Attribute: PRIMARY

Limits: 1 5

Color:

Plot Options - 3D Volume Rendering

x axis Pixel Axis 2 [x]

min/max: 38.2241 ⇌ 1160.78

stretch: 1.00

y axis Pixel Axis 1 [y]

min/max: 38.2241 ⇌ 1160.78

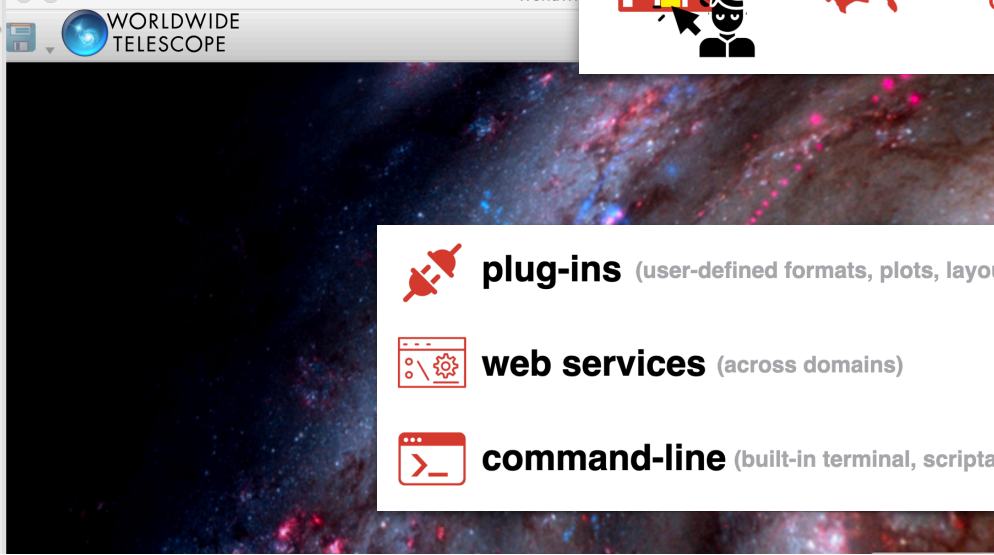
stretch: 1.00




z axis Pixel Axis 0 [z]

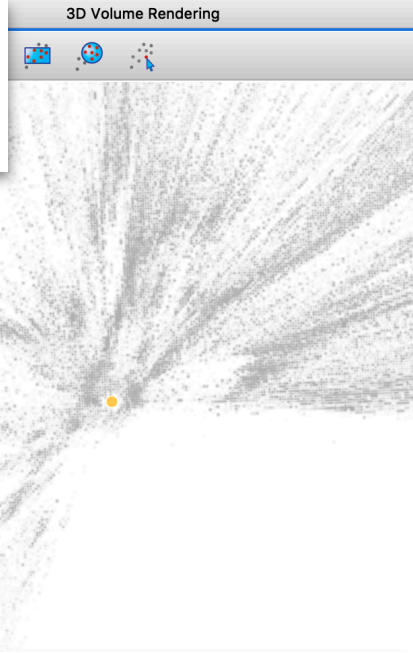
min/max: 5.95402 ⇌ 193.046

stretch: 1.00

reference: Green 2019 3D Dust



-  **plug-ins** (user-defined formats, plots, layouts...)
-  **web services** (across domains)
-  **command-line** (built-in terminal, scriptable)



And how easy/hard is it to use?

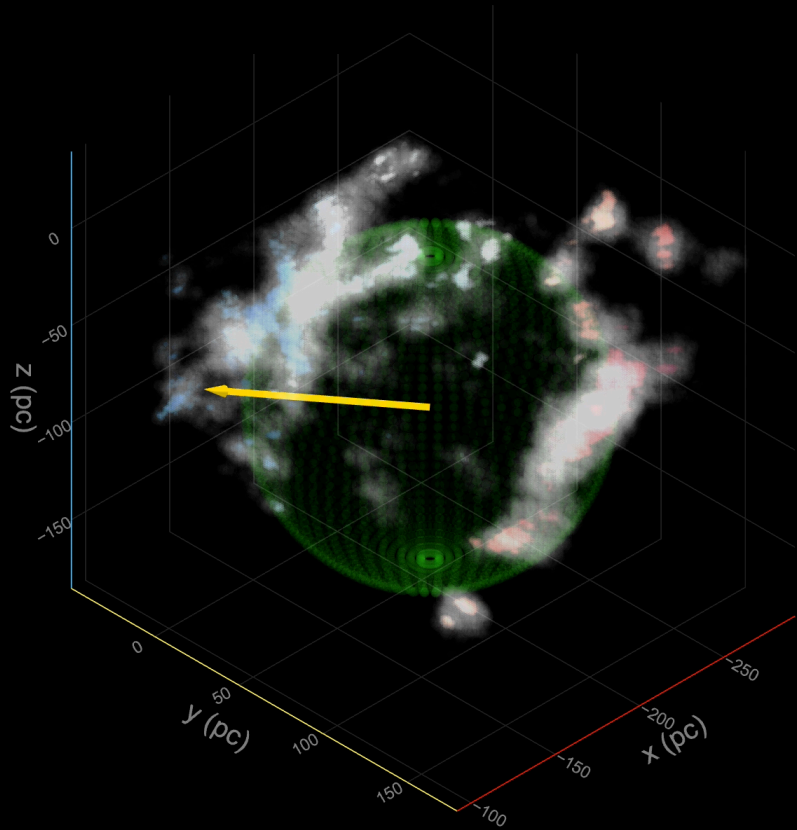
The image is a composite of three main elements:

- Top Left:** A screenshot of a 3D visualization software interface. The central window shows a 3D volume rendering of the Local Bubble, a region of interstellar space. The visualization is a complex, multi-colored structure (blue, purple, yellow, red) with a central core. The interface includes a 'Data Collection' panel on the left with various data sources like 'Local_Axis_FX_Best2018' and 'Local_Axis_FY_Best2018'. Below that is a 'Plot Layers' panel with '3D Volume Rendering' selected. The 'Plot Options' panel at the bottom left shows settings for 'x axis', 'y axis', and 'z axis' with 'Perspective' view selected.
- Top Right:** A snippet of a newspaper article from 'The New York Times' dated January 25, 2022. The headline is 'Local Bubble'. The article discusses the discovery of the Local Bubble, a region of interstellar space. It mentions that the bubble is a vast, empty space that has been carved out of the interstellar medium by the winds of stars. The article also mentions that the bubble is a region of low density and low temperature, and that it is a region of high magnetic field strength. The article is attributed to Alyssa Goodman and João Alves.
- Bottom:** A screenshot of a video call interface. Two participants are visible: Alyssa Goodman and João Alves. The interface includes a 'Mute' button, a 'Stop Video' button, and a 'Leave' button. The participants' names are displayed below their respective video feeds.

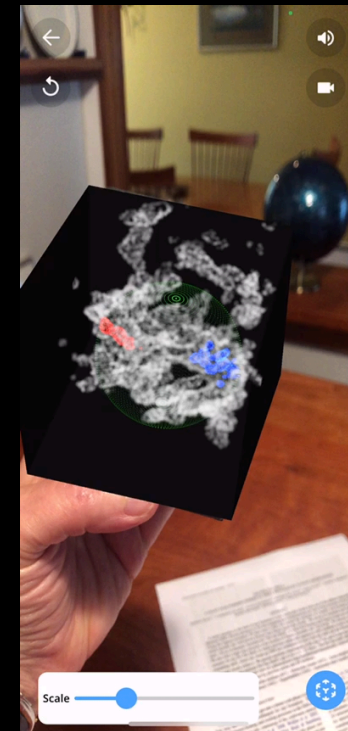
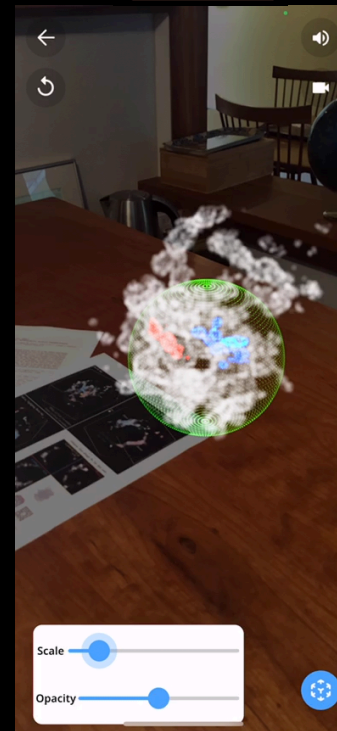
The Future



The Future of Publishing



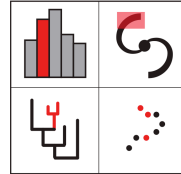
[TINYURL.COM/UNIVERSE-IN-MY-HAND](https://tinyurl.com/universe-in-my-hand)



AUGMENTED REALITY

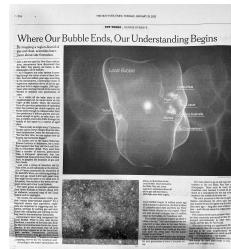
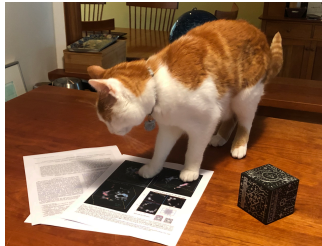
Triumphs of **glue** (and plot.ly) ++

CO SPACES

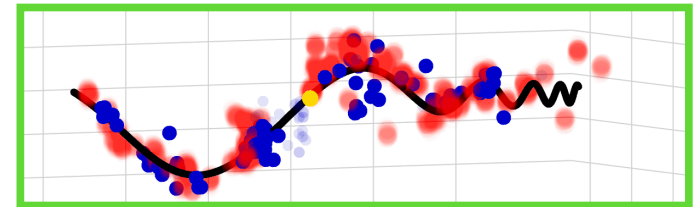


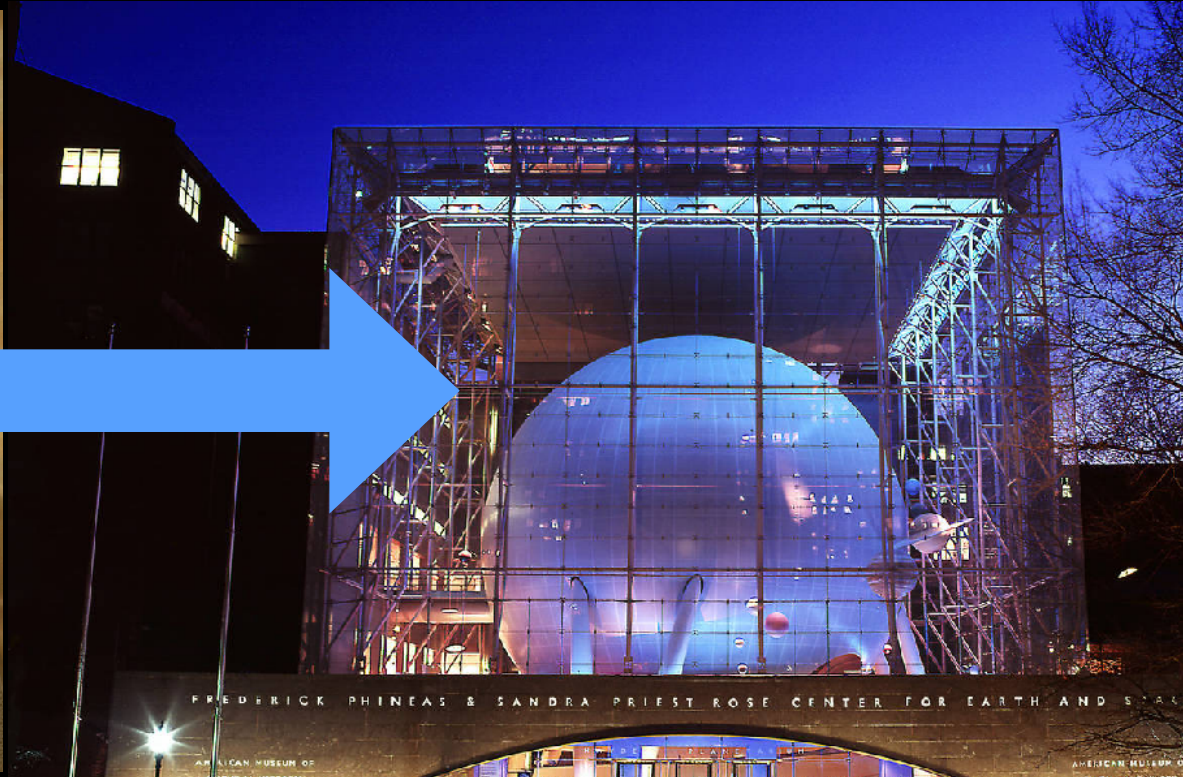
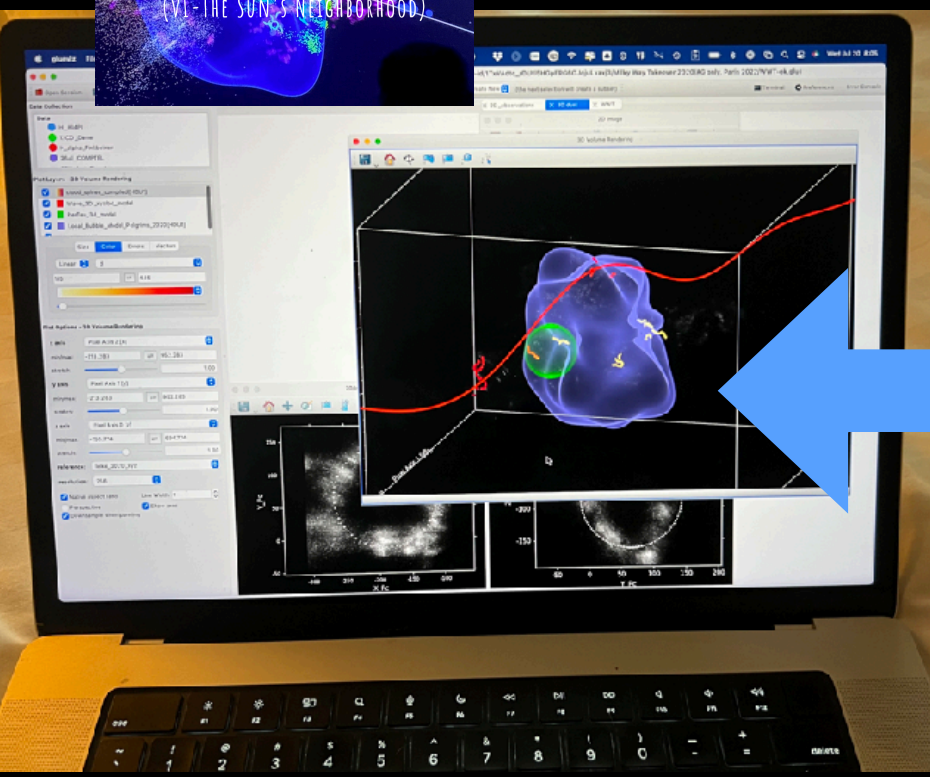
glue
multidimensional data exploration

 **plot.ly**



*This one is EMBARGOED—I will show you a demo, but please **do not post it, or share it.***





A 3D visualization of the Milky Way galaxy, showing a central bulge and spiral arms. The visualization is rendered in a dark blue and purple color scheme, with a prominent orange-red line representing the galactic plane or a specific path. The text 'THE MILKY WAY IN 3D' is overlaid in a white, sans-serif font, with '(V1 - THE SUN'S NEIGHBORHOOD)' in a smaller font below it.

THE MILKY WAY IN 3D

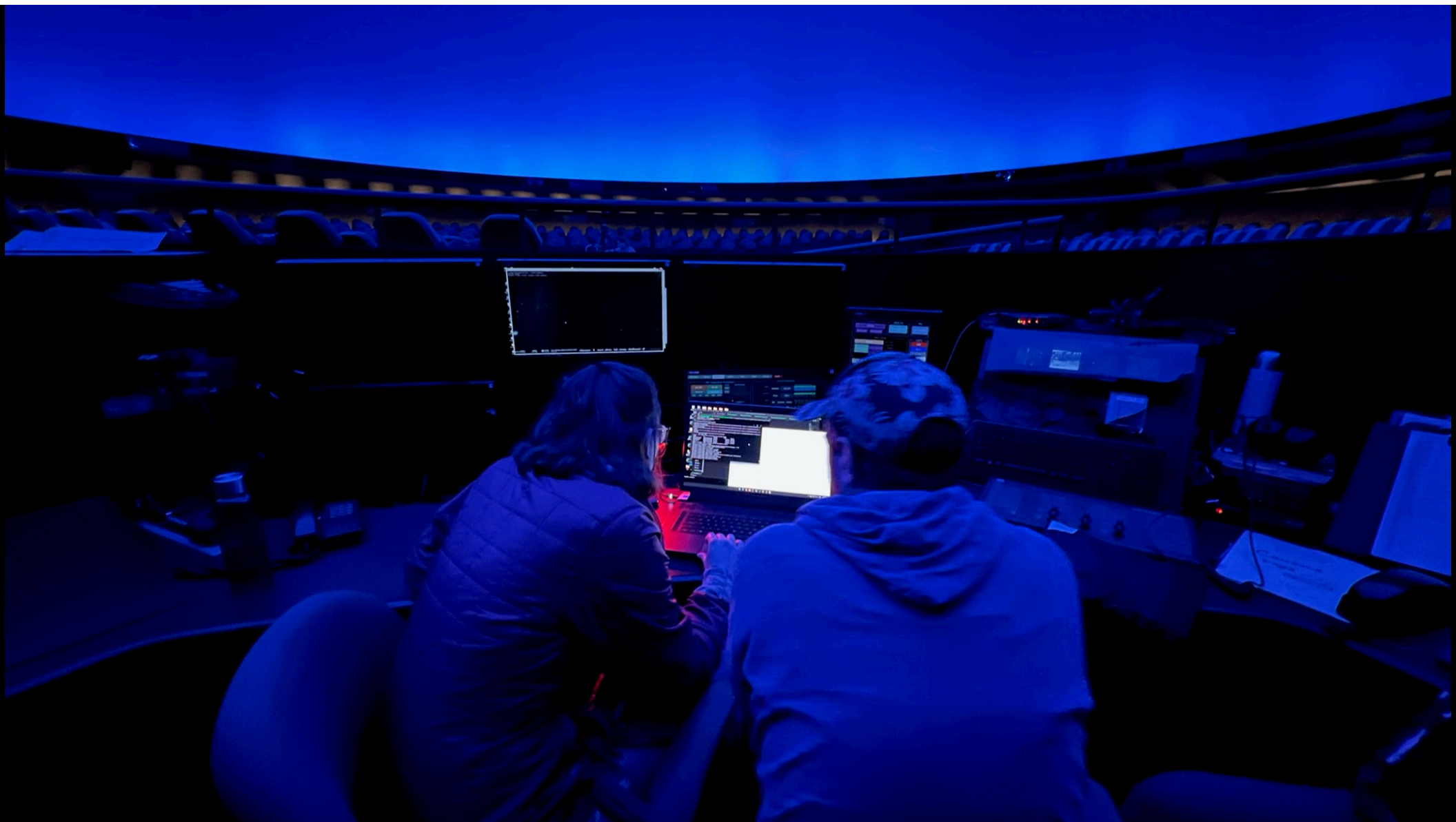
(V1 - THE SUN'S NEIGHBORHOOD)

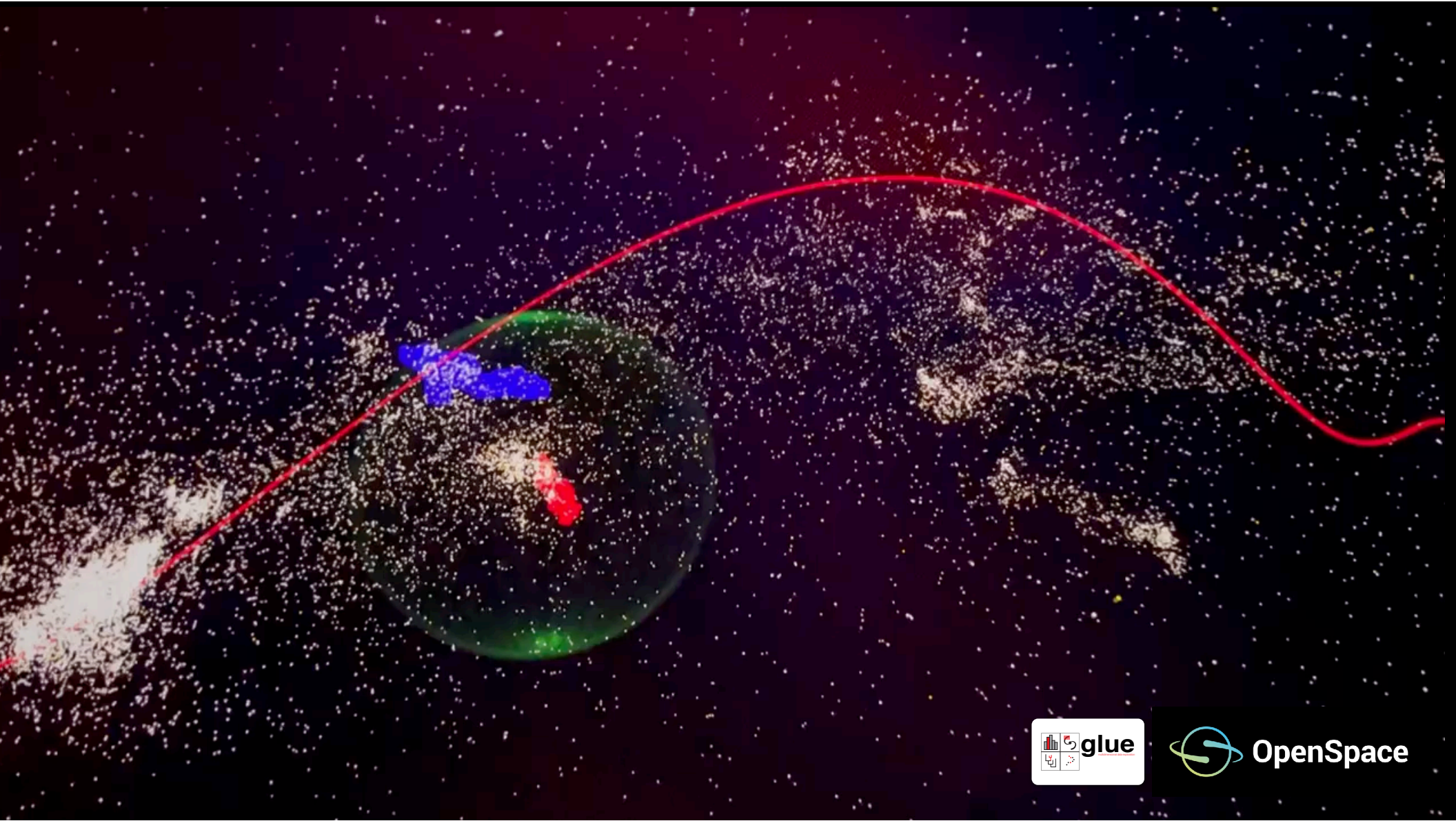
Welcome to a new view of the Milky Way... in 3D!

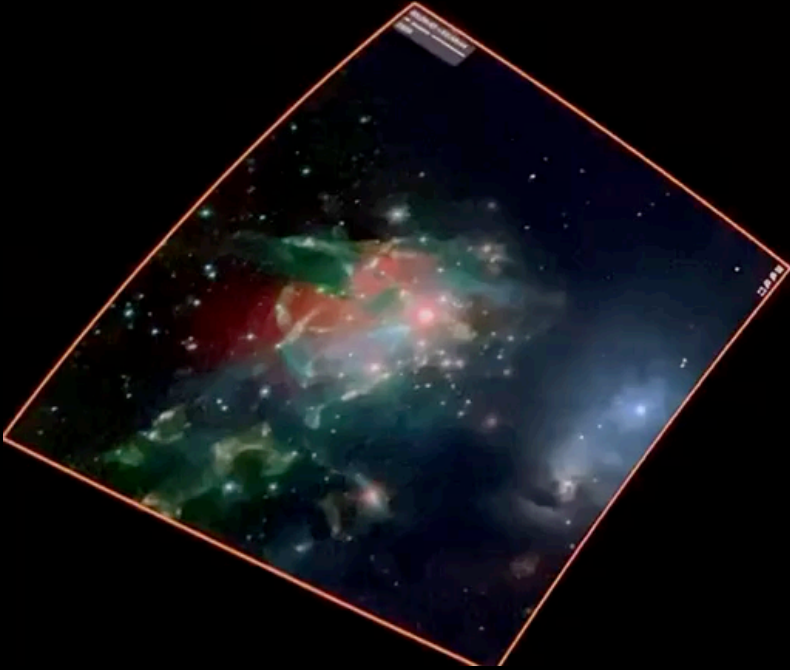
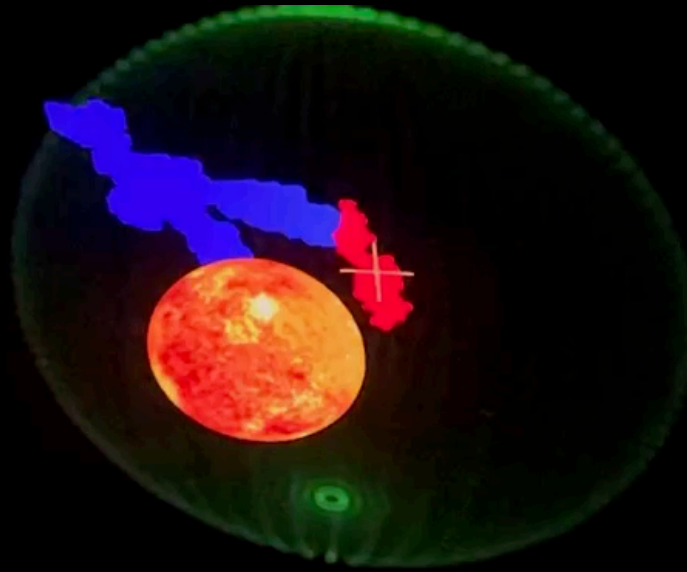
Soon, milkyway3d.org will serve as a hub for the interconnected set of outreach, education, and research resources that will result from the interconnections we're in the process of making.

Our project includes new software development; approaches to data sharing; and scientific research questions propelling our collaboration forward.

milkyway3d.org



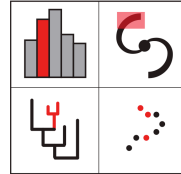






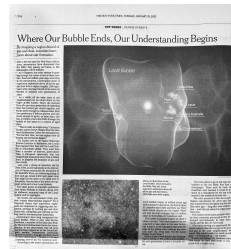
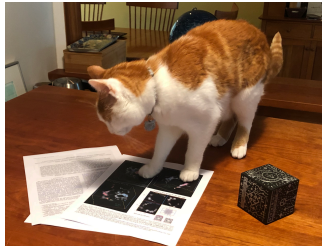
Triumphs of glue (and plot.ly) ++

CO SPACES

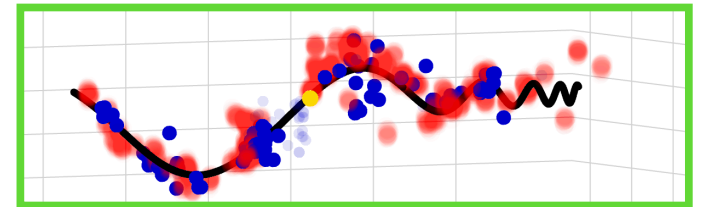


glue
multidimensional data exploration

plotly



This one is EMBARGOED—I will show you a demo, but please do not post it, or share it.



Progeny of glue

Progeny of glue

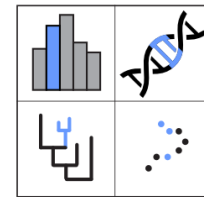


later...a Google Drive souvenir...
with plenty of notebook samples.





glue
solutions
inc.



glue
genes

“High-dimensional Data”: Genomics

Data Collection

Data

- chr3_3DGNOME
- ChIA-PET_Rep1_coverage_ENCFF310
- ChIA-PET_Rep1_loops_ENCFF310TX
- ChIA-PET_Rep1_peaks_ENCFF578RML

Subsets

- Isolated Region

Plot Layers

- ChIA-PET_Rep1_loops_ENCFF310TX
- ChIA-PET_Rep1_peaks_ENCFF578RML
- ChIA-PET_Rep1_coverage_ENCFF310

color

opacity

Plot Options - Genome Track Viewer

Limits Axes Legend

Chrom

Range

Loops

Show Genes

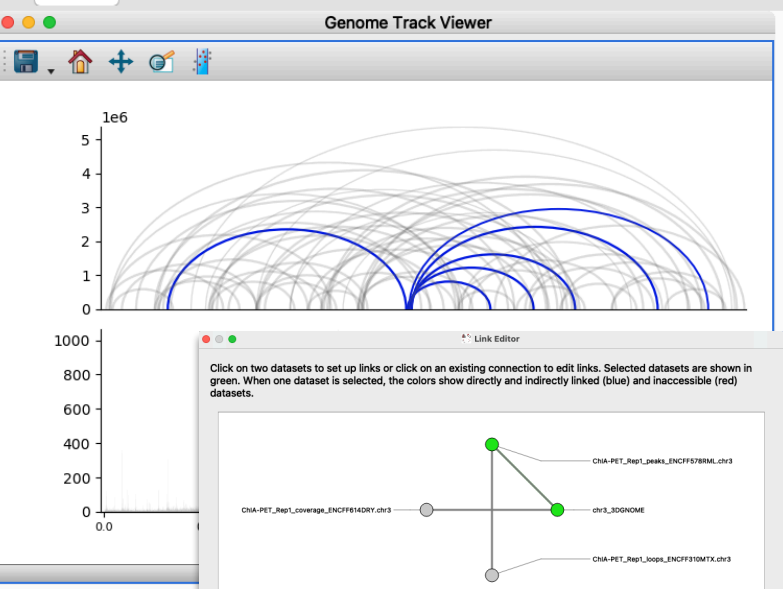
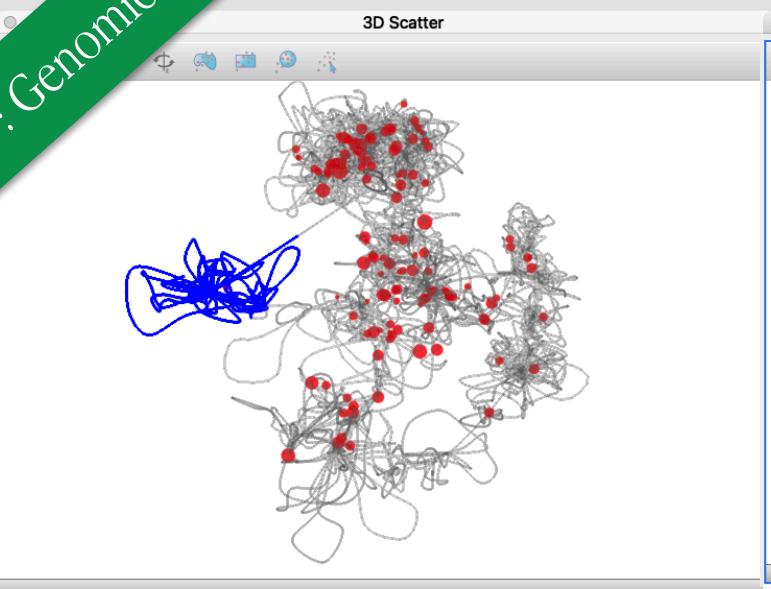


Table: ChIA-PET_Rep1_peaks_ENCFF578RML.chr3

	chr	end	start	peak_intensity	x	y
26	chr3	5027016	5026505	4.87089987...	-16.95555...	-20.22317..
27	chr3	5027804	5027264	4.79005547...	-15.96221...	-20.38275.
28	chr3	5058836	5057743	4.06550913...	-29.73126...	-18.92818..
29	chr3	5121954	5121471	3.4065460...	-12.510781...	-14.179771.
30	chr3	5122810	5122272	4.781716172...	-12.42688...	-10.85097.
31	chr3	5123968	5123002	3.9530344...	-12.511697...	-7.803144.
32	chr3	5137283	5136938	2.99992079...	-24.46438...	-5.097149.
33	chr3	5426373	5425735	3.2009293...	-41.73639...	33.330951
34	chr3	5427401	5426635	3.73743405...	-41.83707...	32.817504.
35	chr3	6994724	6994337	4.35688817...	9.6423483...	-7.427244.

Link Editor

Click on two datasets to set up links or click on an existing connection to edit links. Selected datasets are shown in green. When one dataset is selected, the colors show directly and indirectly linked (blue) and inaccessible (red) datasets.

Dataset 1: ChIA-PET_Rep1_peaks_ENCFF578RML.chr3

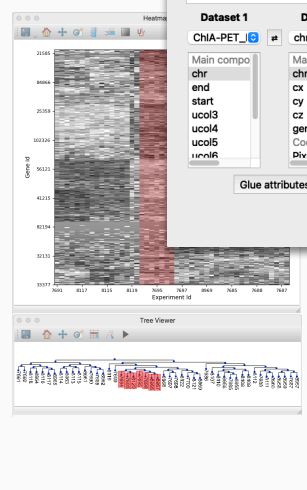
Dataset 2: chr3_3DGNOME

Links between Dataset 1 and Dataset 2: identity(start <-> genome_position), identity(chr <-> chr)

Link details

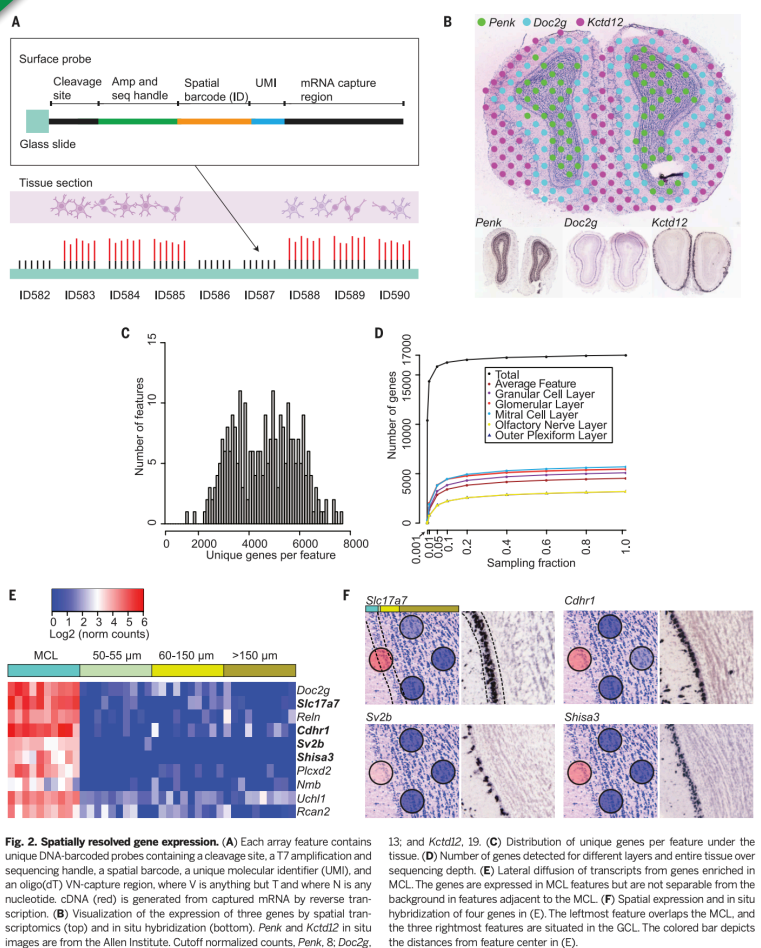
Link components: Dataset 1 attributes (x: chr), Dataset 2 attributes (y: chr)

Buttons: Glue attributes, Create advanced link, Remove link, Cancel, OK

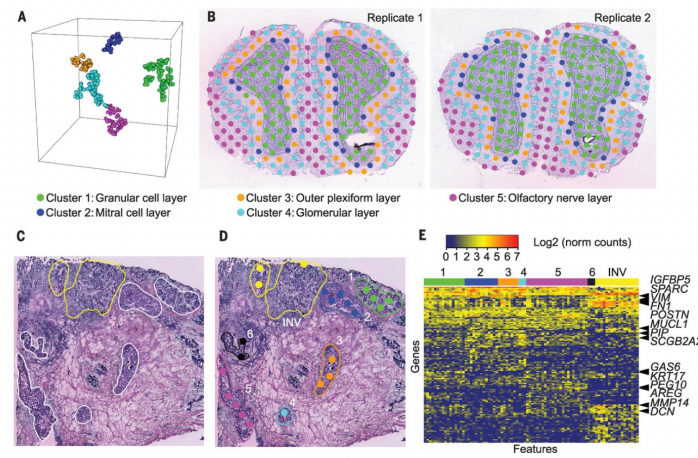


“High-dimensional Data”: Genomics
 Visualization and analysis of gene expression in tissue sections by spatial transcriptomics

Ståhl PL,^{1,2*} Salmén F,^{2*} Vickovic S,^{2*} Lundmark A,^{2*} Navarro JF,^{2*} Magnusson J,^{2*} Giacomello S,^{2*} Asp M,^{2*} Westholm JO,^{2*} Huss M,^{2*} Mollbrink A,^{2*} Linnarsson S,^{2*} Codeluppi S,^{2*} Borg Å,^{2*} Pontén F,^{2*} Costea PI,^{2*} Sahlén P,^{2*} Mulder J,^{2*} Bergmann O,^{2*} Lundeberg J,^{2*} Frisén J (2016) Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*, 353(6294):78–82. <https://doi.org/10.1126/science.aaf2403>



Spatial Transcriptomics



Ståhl PL, Salmén F, Vickovic S, Lundmark A, Navarro JF, Magnusson J, Giacomello S, Asp M, Westholm JO, Huss M, Mollbrink A, Linnarsson S, Codeluppi S, Borg Å, Pontén F, Costea PI, Sahlén P, Mulder J, Bergmann O, Lundeberg J, Frisén J (2016) Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*, 353(6294):78–82. <https://doi.org/10.1126/science.aaf2403>

High-dimensional Data: Astronomy

High-dimensional Data: Genomics ++

The image displays a collection of software windows used for high-dimensional data analysis. On the left, under the 'Astronomy' header, there are several 3D scatter plots and a 'Replicate 2' visualization showing a cluster of points. The right side, under the 'Genomics ++' header, features a '3D Scatter' plot of genomic data, a 'Genome Track Viewer' showing arcs and tracks, and a table of ChIA-PET peaks. A histology image is also visible in the bottom left corner.

chr	y	end	start	peak_intensity	x	y
26	chr3	5027016	5026505	4.87089987...	-16.95555...	-20.22317.
27	chr3	5027804	5027264	4.79005547...	-15.96221...	-20.38275.
28	chr3	5058836	5057743	4.06550913...	-29.73126...	-18.92818.
29	chr3	5121954	5121471	3.4065460...	-12.510781...	-14.179771
30	chr3	5122810	5122272	4.781716172...	-12.42688...	-10.85097.
31	chr3	5123968	5123002	3.9530344...	-12.511697...	-7.803144.
32	chr3	5137283	5136938	2.99992079...	-24.46438...	-6.097149.
33	chr3	5426373	5425735	3.2009293...	-41.73639...	33.330951
34	chr3	5427401	5426635	3.73743405...	-41.83707...	32.817504.
35	chr3	6994724	6994337	4.35668817...	9.6423483...	-7.427244.

2023 features of gluegenes

Reads **Genomic Data**

Bed
Bedgraph
Bedpe
BigWig
RNA-seq and ATAC-seq data matrices
3D models from 3D-GNOME
Single-cell data (AnnData)
CSV
Excel
Numpy Savefile
HDF5
Images

Standard and **Custom Viewers**

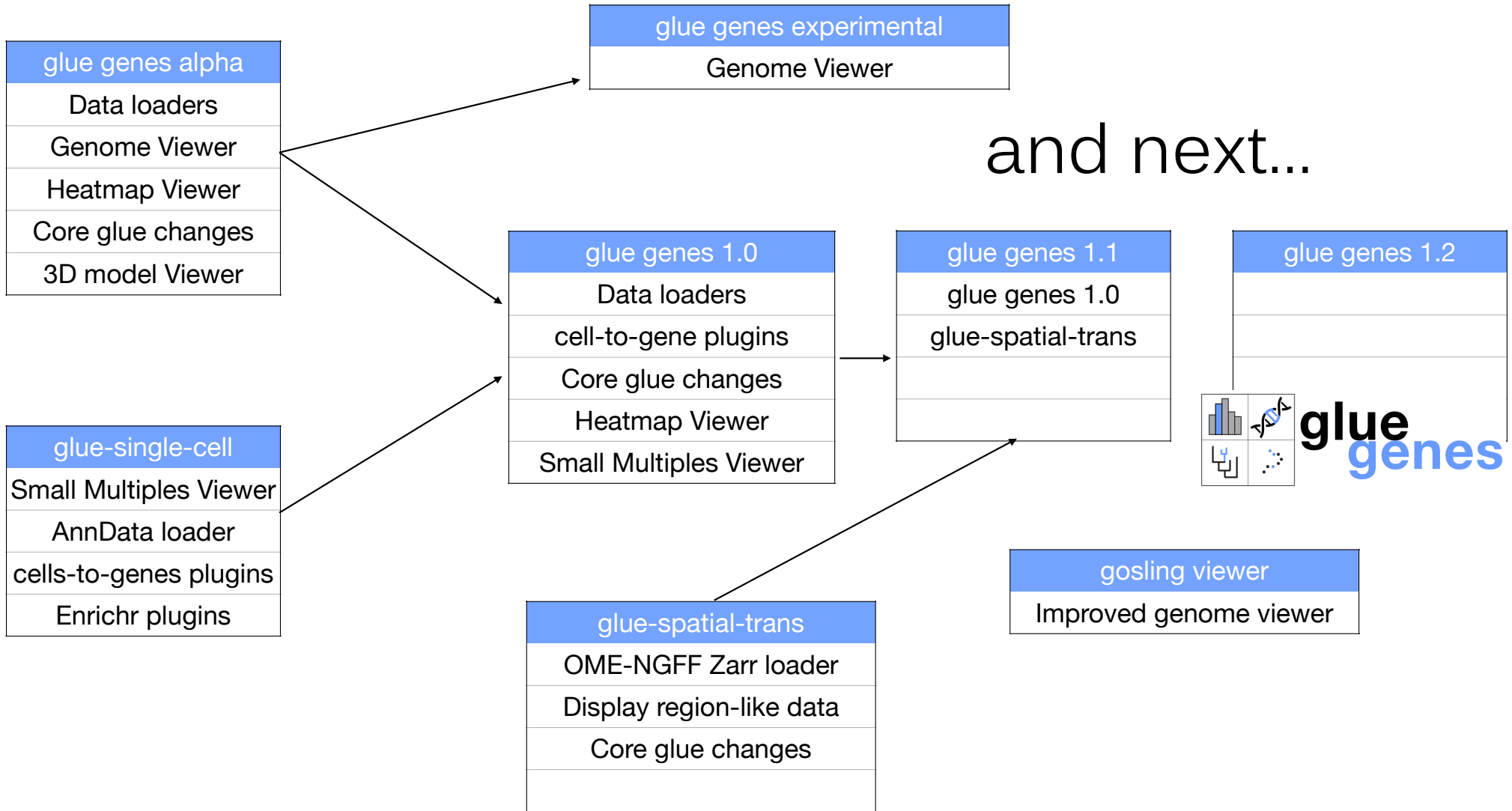
1D histogram
2D scatterplot
2D images
3D scatter
3D volume
2D heatmap
QTL viewer
Small multiples viewer

Analysis plug-ins

Get differentially expressed genes from two subsets of cells
Measure/display expression of gene subsets over cells
Get KEGG pathways for gene subsets



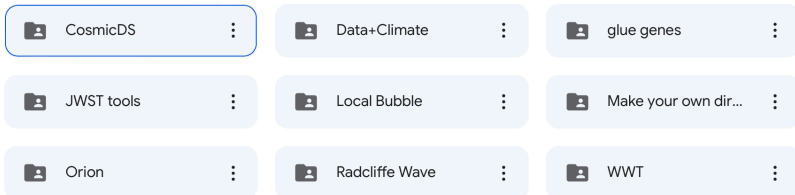
Customized for Genomics



Progeny of glue



later...a Google Drive souvenir...
with plenty of notebook samples.

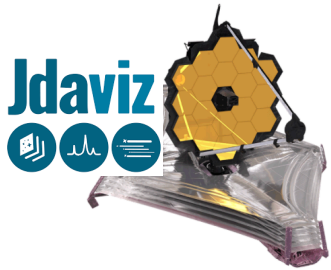


glupyter (a.k.a. "glue jupyter")

gluesolutions.io/
the-software/
glupyter

"glupyter" is a union of [glue](#) and [Jupyter](#) software environments. We think it may well be the future of glue, or "glue-qt" as experts sometimes call the desktop app version of glue. This webpage, hosted openly and freely by [glue solutions, inc.](#), serves as a clearinghouse for current information about open-source glupyter-related projects. Some of these projects are funded by government agencies (notably [NSF](#) and [NASA](#)), others by private foundations (e.g. [The Gordon and Betty Moore Foundation](#)), some as part of corporate collaborations (e.g. [Harvard+Google Data+Climate](#)), and some by open-source consulting work carried out by [glue solutions, inc.](#)

The [glue-jupyter](#) GitHub repository is fully open,
and more detail can be found on this [Read the Docs](#) page.



Quick insights for Images,
Spectra

JDAViz

includes: ImViz, CubeViz, SpecViz,
MOSViz

Sponsor: NASA, James Webb Space
Telescope

[Read more \(blog post at 10QViz.org\)...](#)



Open-Source GIS Data
Exploration

SAVE

Search-Analysis-Visualization-
Environment

Sponsors: Harvard+Google
Data+Climate

[Read more
at Data+Climate site...](#)



Data Science Education

Cosmic Data Stories

Sponsor: NASA, Science Activation
Program (funded proposal)

[Read more
at CosmicDS website...](#)

[GitHub](#)

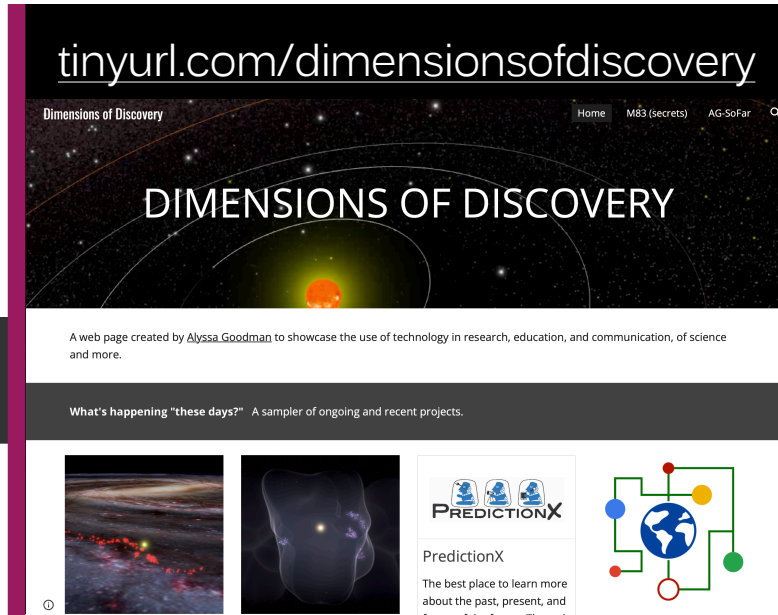


bringing glue to
JupyterLab

glupyter prototype

Sponsors: The Gordon and Betty
Moore Foundation and the National
Science Foundation

[Read more in the justification of the
GBF proposal, awarded to
Harvard, and watch this 2022 demo
video](#)



"LEVELS" of interaction/users



Guided experience for
learners, no coding, only
interactive webpages



dashboard-style




Fully flexible,
scriptable, extensible.



Coming in 2023, glue in  v.0.1



GORDON AND BETTY
MOORE
FOUNDATION

grant to: glue (A. Goodman),
Jupyter (F. Perez) developed in
collaboration with S. Corlay et al at 

File Edit View Run Kernel Tabs Settings Help

Filter files by name

Name	Last modified
3D Objects	7 months ago
anaconda3	3 months ago
ansel	7 months ago
Contacts	7 months ago
Documents	7 months ago
Downloads	7 days ago
Favorites	7 months ago
Links	7 months ago
Music	7 months ago
OneDrive	5 minutes ago
Python 3.10	6 months ago
Saved Games	7 months ago
scikit_learn_data	2 months ago
Searches	7 months ago
Videos	8 minutes ago
2016_Building_Energy_B...	7 days ago
bag_of_words.csv	2 months ago
data.csv	6 months ago
data34.csv	6 months ago
EdStatsData.csv	3 months ago
Espace.py	6 months ago
ExtractionWebUK.py	6 months ago
Fonctions.ipynb	a month ago
fropenfoodfacts.org.pro...	2 months ago
functions.py	a month ago
hello.txt	6 months ago
house.csv	2 months ago
my_courses.csv	2 months ago
my_course2.csv	2 months ago
mystery.csv	2 months ago
Nettoyage de Jeu.ipynb	3 months ago
P3_kNN_reconnaissance...	2 months ago
PackageCSVexo	6 months ago
personnes.csv	3 months ago
Projet 2.ipynb	3 months ago
Projet 2NC.ipynb	2 months ago
Projet 3 Brouillon2.ipynb	a month ago

Launcher

Notebook

Python 3 (ipykernel)

Console

Python 3 (ipykernel)

Other

Terminal Text File Markdown File Python File Show Contextual Help


Glue Visualization

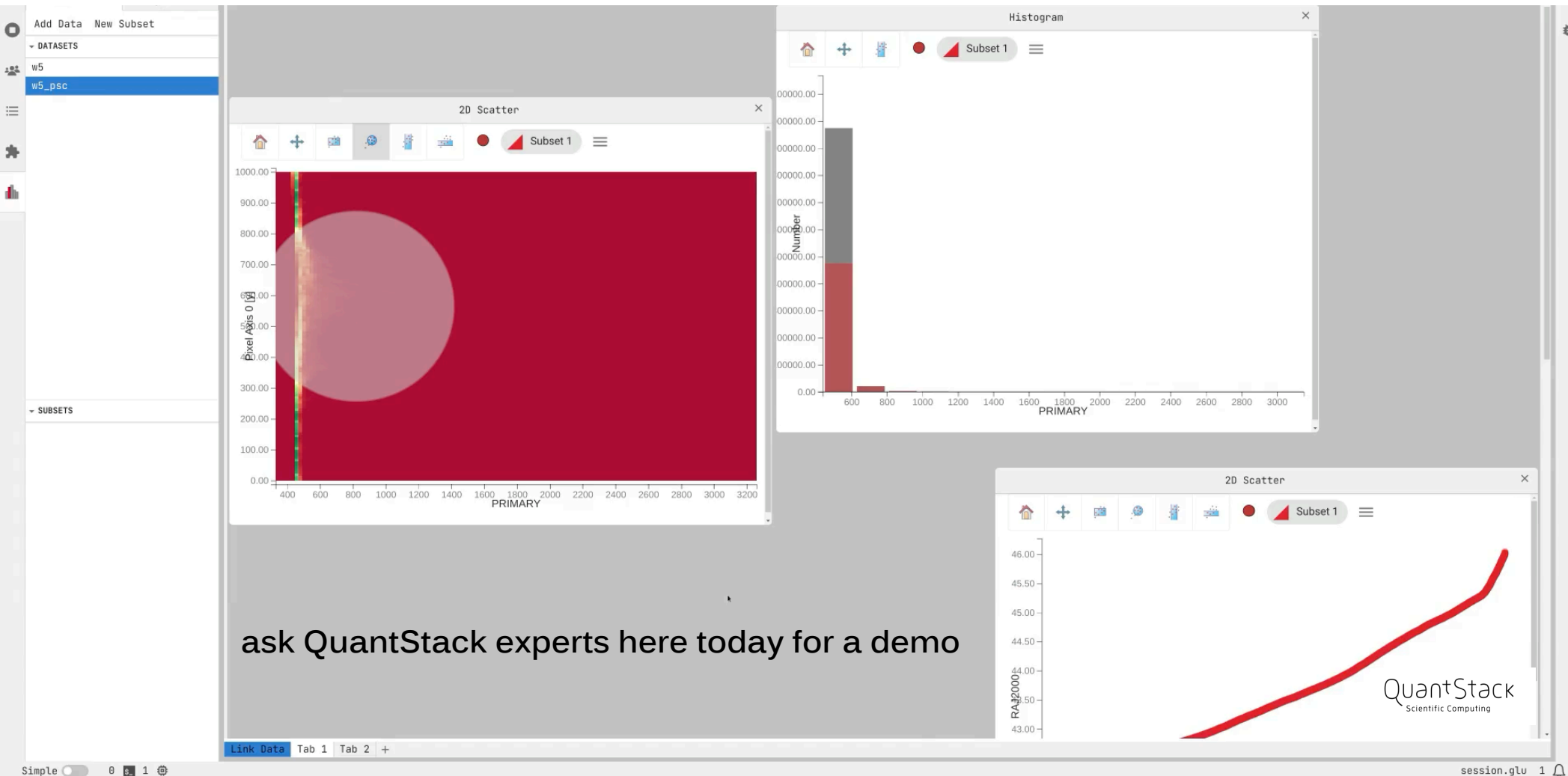
New Session

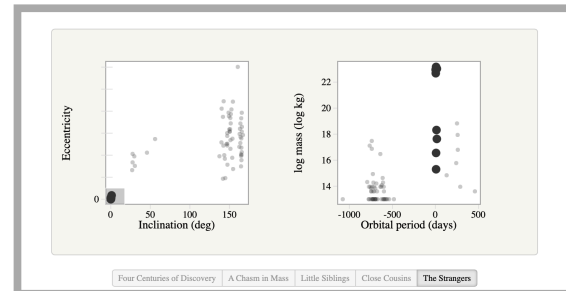
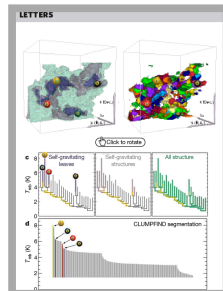
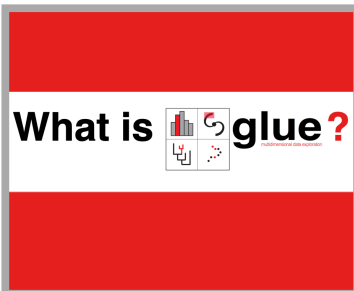
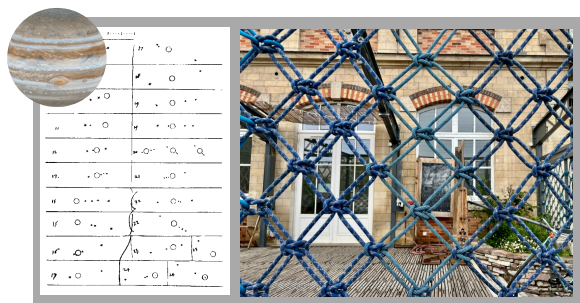
video & UX design: Gabriela Vives

Coming in 2023, glue in  v.0.1

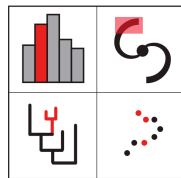


grant to: glue (A. Goodman),
Jupyter (F. Perez) developed in
collaboration with S. Corlay et al at 





CO SPACES



glue
multidimensional data exploration

 **plotly**



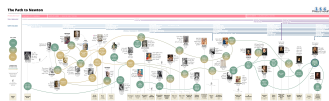
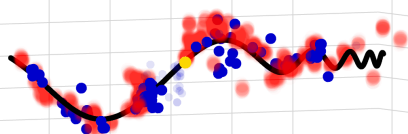
 **gluepyter**
multidimensional data exploration

 **glue genes**


CosmicDS



Jdaviz

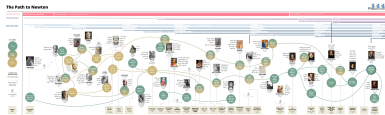
The **TIMELINE**
CONSORTIUM

 **astrophysics**
data system

 **OpenAI**

1  **ViZ**

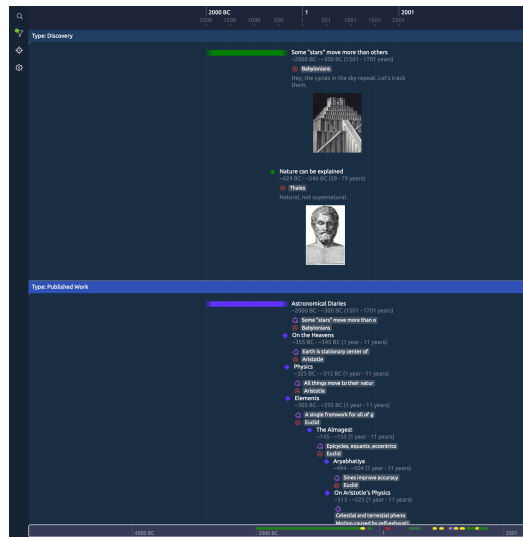
Seeds for discussion...



The TIMELINE
CONSORTIUM



Seeds for discussion...



Forging New Horizons in Astronomy Research with Foundation Models (a proposal SELECTED by Microsoft, April 26, 2023)

Principal Investigator

Alyssa Goodman (Center for Astrophysics | Harvard & Smithsonian)

Co-Investigators

Ioana (Jo) Clucă & Yuan-Sen Ting (Australian National University)

Alberto Accomazzi (NASA ADS & Center for Astrophysics | Harvard & Smithsonian)

Josh Peek (Space Telescope Science Institute)

Introduction

Dating back nearly five thousand years, Astronomy has aimed to unravel the mysteries of the Universe and push the boundaries of theoretical understanding. The advent of the Internet has allowed astronomers to archive our corpus of knowledge, with NASA's Astrophysics Data System* (ui.adsabs.harvard.edu) hosting over 15 million resources, representing essentially all of the astronomical literature used by researchers (Accomazzi et al. 2015, Borgman & Wofford 2021). Recent breakthroughs in large language models (LLMs, e.g., Vaswani et al. 2017; Devlin et al. 2018; Brown et al. 2020) are now empowering researchers to draw insights from the expansive and intricate body of astronomy literature. For example, the LLMs can help researchers navigate the various scientific interpretations of research findings, which may disagree in non-trivial ways. Moreover, the multidisciplinary aspect of Astronomy enables the utilization of LLMs to reveal hidden relationships within our knowledge corpus, thereby opening the possibility of formulating novel scientific hypotheses.

Goals

The primary goal of this project is to improve human interaction with astronomy literature by leveraging the capabilities of foundation models such as the SoTA GPT-4 LLM (OpenAI 2023) in a focused manner. To this end, we will collaborate closely with NASA ADS to define a set of astronomy-specific downstream tasks (ADTs) such as extracting questions, answering with references, comparative analyses between different research papers, creating mind maps, scientific summarization, and generating new scientific ideas. We will then curate an evaluation dataset representative of a domain in Astrophysics (e.g., *Galaxy Evolution*, *Galactic Astronomy*) on which to evaluate the LLMs' performance.

The project aims to: a) develop a targeted approach for engaging GPT-4 with complex astronomy literature using in-context prompting; b) evaluate the performance of GPT-4 across astronomy-specific tasks (ADTs); c) create input-output pairs to cover the domain tasks; d) use the GPT-4 generated input-output pairs to fine-tune LLMs available through the Azure OpenAI Service; e) explore the potential of foundation models for generating new scientific hypotheses.

Scope of the research

- In-Context Prompting Optimization for GPT-4.
 - Employ in-context prompting techniques using tools such as LangChain (Chase 2023) to develop a practical approach for providing GPT-4 with a broad context of astronomy literature, expanding on the approach proposed in Clucă & Ting 2023.

≡ MENU



TEN QUESTIONS TO ASK WHEN CREATING A VISUALIZATION

The 10 Questions

- Who** | Who is your audience? How expert will they be about the subject and/or display conventions?
- Explore-Explain** | Is your goal to explore, document, or explain your data or ideas, or a combination of these?
- Categories** | Do you want to show or explore pre-existing, known, human-interpretable, categories?
- Patterns** | Do you want to identify new, previously unknown or undefined patterns?
- Predictions & Uncertainty** | Are you making a comparison between data and/or predictions? Is representing uncertainty a concern?
- Dimensions** | What is the intrinsic number of dimensions (not necessarily spatial) in your data, and how many do you want to show at once?
- Abstraction & Accuracy** | Do you need to show all the data, or is summary or abstraction OK?
- Context & Scale** | Can you, and do you want to, put the data into a standard frame of reference, coordinate system, or show scale(s)?
- Metadata** | Do you need to display or link to non-quantitative metadata? (including captions, labels, etc.)
- Display Modes** | What display modes might be used in experiencing your display?



Now, visit the 10QViz conversation! There's so much more to talk about.

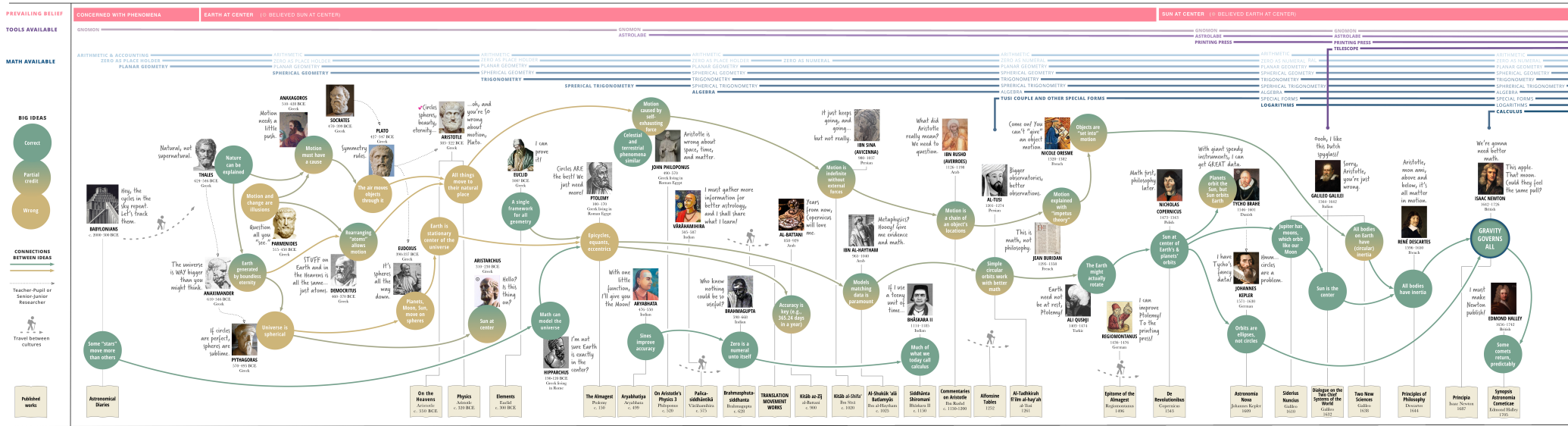


Curious about the **origins** of 10QViz? Try the [About](#) page.
Want to learn **how best to use and participate** in 10QViz? Try the [How to](#) page.
Want to read about the **scholarship** behind 10QViz.org's questions?
And, there's more at our **YouTube channel!**
Write to ask for a draft of our research paper, Coltekin & Goodman 2019.

The Path to Newton



The Path to Newton



© Harvard University, created by Alyssa Goodman, Jodi Brodinsky, Drew Lichtenthaler & Katie Peck, re-use is allowed with attribution, version 1, 2018

demo: path-to.org

or



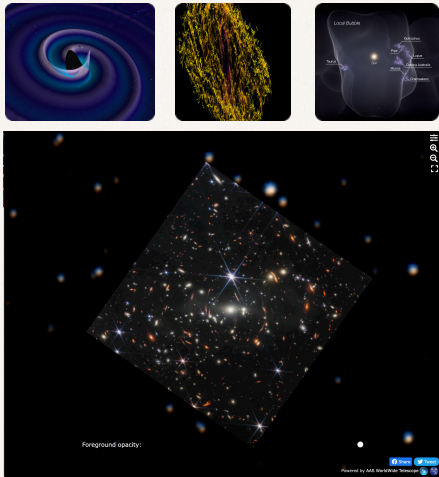
Opinion

The New Universe

MEMPHIS, SUNDAY OCTOBER 23, 2022

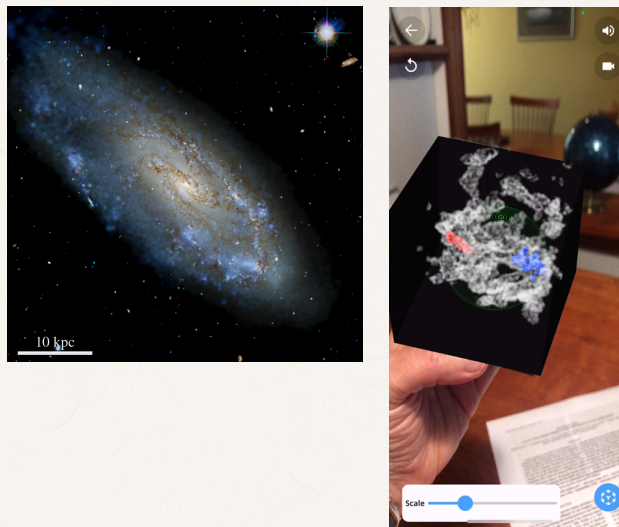
WHAT DO EXPENSIVE NEW TELESCOPES DO FOR HUMANITY TODAY?

Are mega-projects like ALMA, LIGO, JWST, and Gaia worth the billions?



ARE COMPUTERS THE NEW TELESCOPES?

New galaxies in-silico, the early Universe without physics, and new stars forming in your hand.



IS ASTROPHYSICS BEING (RE)ORGANIZED?

Lone stargazers are a rarer and rarer breed in professional astronomy. Teams and data scientists seem the way of the future, and tools that talk to each other are essential.



**ARE COMPUTERS THE
NEW TELESCOPES?**

The Path to Newton

PREVAILING BELIEF

CONCERNED WITH PHENOMENA | EARTH AT CENTER (☉ BELIEVED SUN AT CENTER)

TOOLS AVAILABLE

GNOMON

GNOMON
ASTROLABE

MATH AVAILABLE

ARITHMETIC & ACCOUNTING | ARITHMETIC | ARITHMETIC | ARITHMETIC
 ZERO AS PLACE HOLDER | ZERO AS PLACE HOLDER | ZERO AS PLACE HOLDER | ZERO AS NUMERAL
 PLANAR GEOMETRY | PLANAR GEOMETRY | PLANAR GEOMETRY | PLANAR GEOMETRY
 SPHERICAL GEOMETRY | SPHERICAL GEOMETRY | SPHERICAL GEOMETRY | SPHERICAL GEOMETRY
 TRIGONOMETRY | SPHERICAL TRIGONOMETRY | SPHERICAL TRIGONOMETRY | ALGEBRA

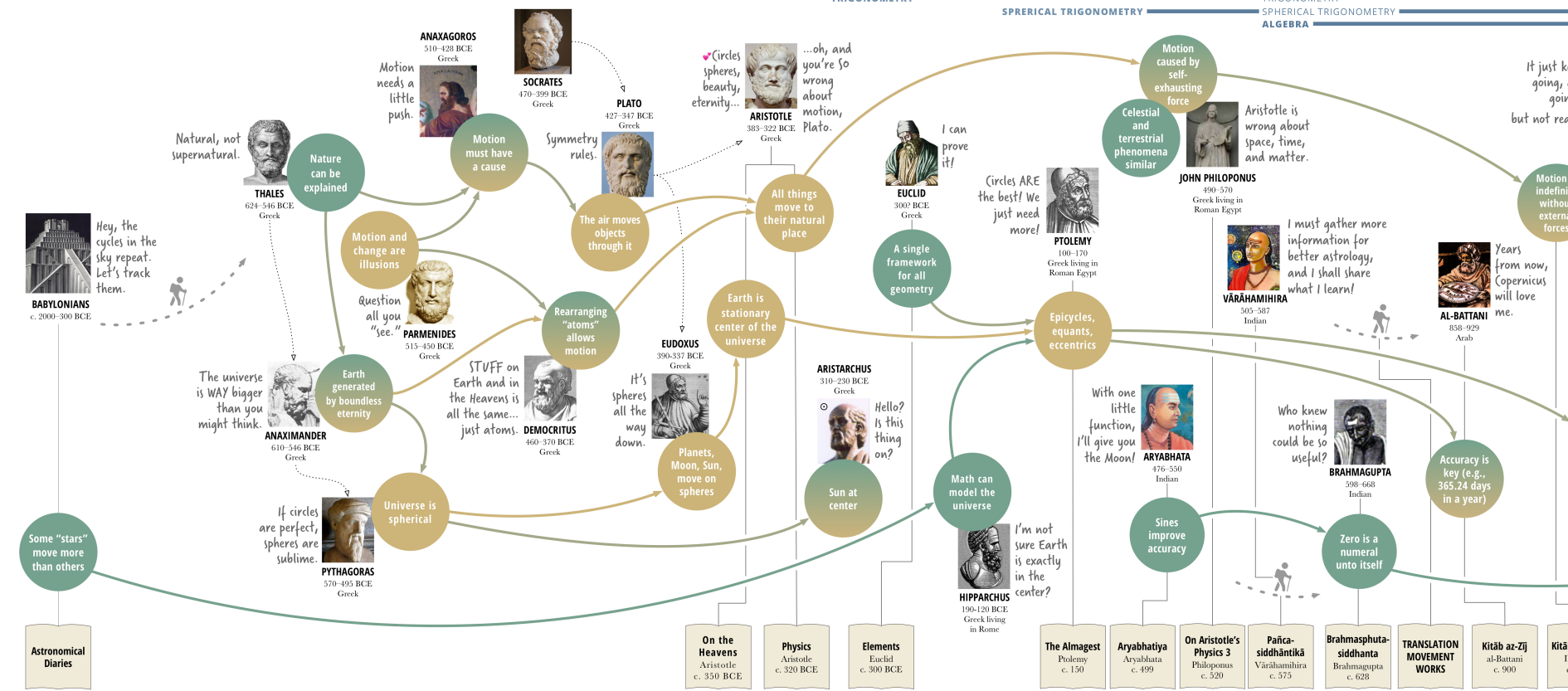
- BIG IDEAS**
- Correct
 - Partial credit
 - Wrong

CONNECTIONS BETWEEN IDEAS

Teacher-Pupil or Senior-Junior Researcher

Travel between cultures

Published works



GNOMON
ASTROLABE

GNOMON
ASTROLABE
PRINTING PRESS

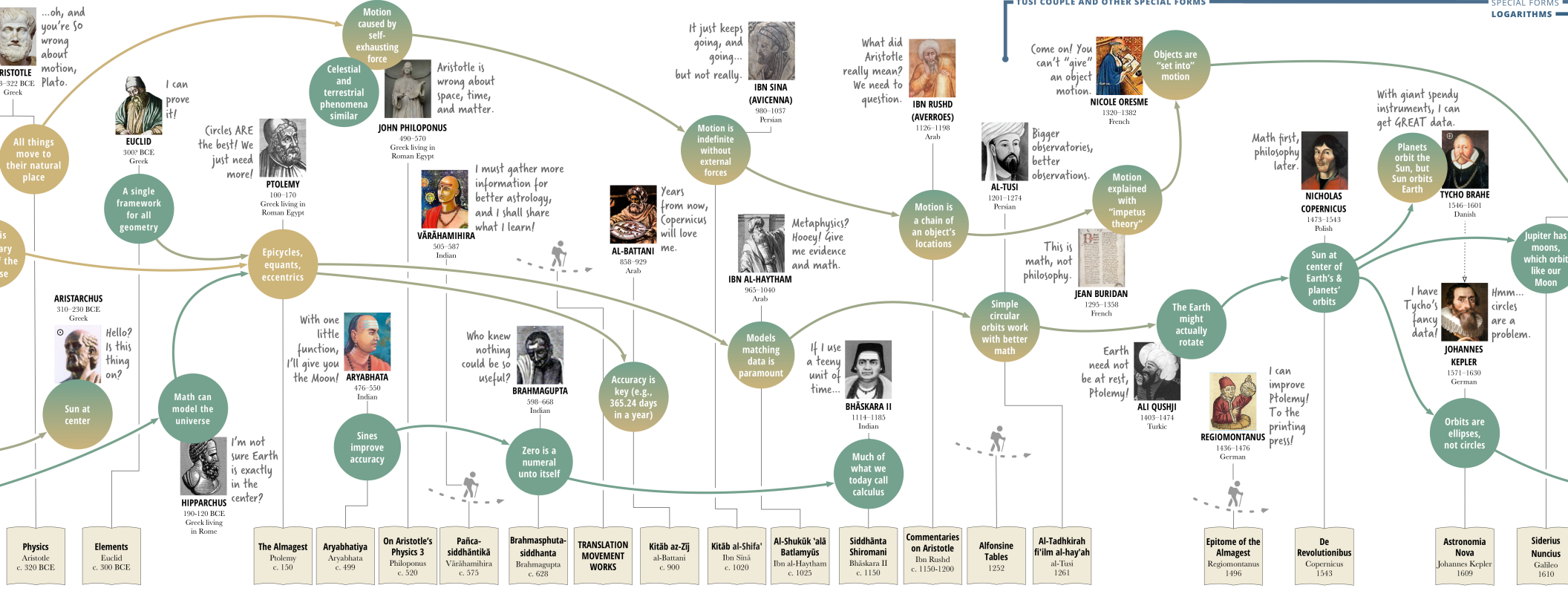
ARITHMETIC
ZERO AS PLACE HOLDER
PLANAR GEOMETRY
SPHERICAL GEOMETRY
TRIGONOMETRY

ARITHMETIC
ZERO AS PLACE HOLDER
ZERO AS NUMERAL
PLANAR GEOMETRY
SPHERICAL GEOMETRY
TRIGONOMETRY
SPHERICAL TRIGONOMETRY
ALGEBRA

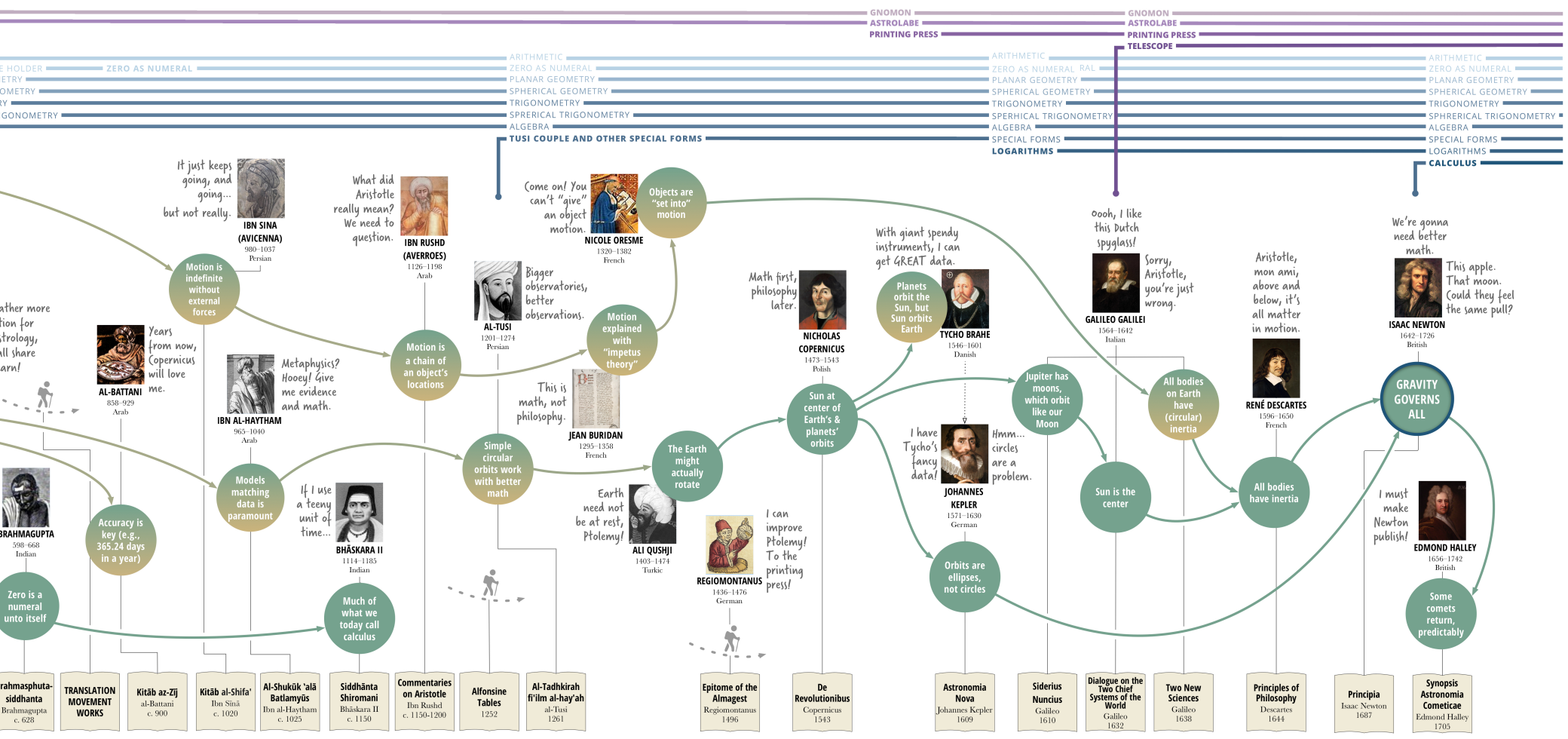
ARITHMETIC
ZERO AS NUMERAL
PLANAR GEOMETRY
SPHERICAL GEOMETRY
TRIGONOMETRY
SPHERICAL TRIGONOMETRY
ALGEBRA

ARITHMETIC
ZERO AS NUMERAL
PLANAR GEOMETRY
SPHERICAL GEOMETRY
TRIGONOMETRY
SPHERICAL TRIGONOMETRY
ALGEBRA
SPECIAL FORMS
LOGARITHMS

SPHERICAL TRIGONOMETRY

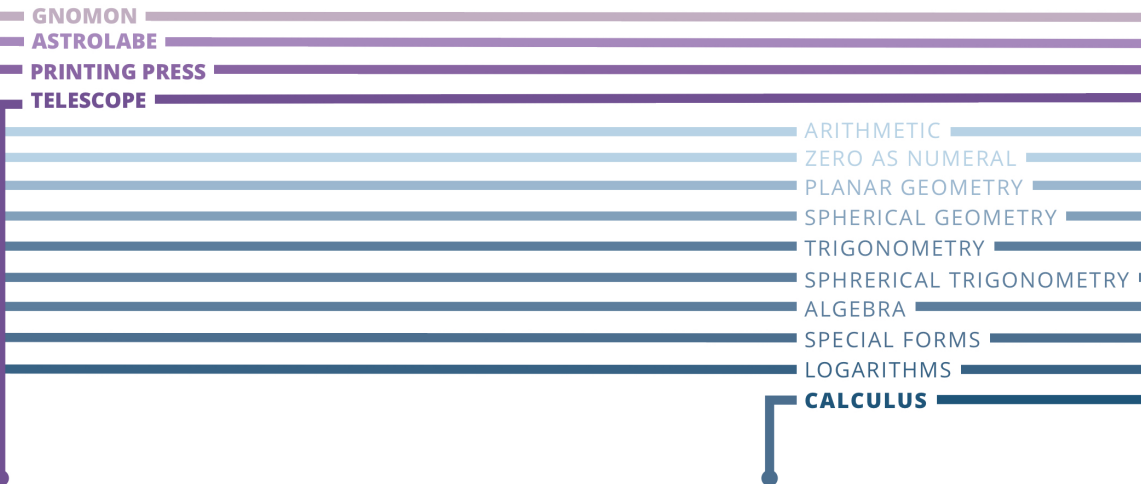


SUN AT CENTER (vs BELIEVED EARTH AT CENTER)





“ARE COMPUTERS THE NEW TELESCOPES?”



ELECTRONIC COMPUTERS

NUMERICAL SIMULATION

BAYESIAN STATISTICS

INTERACTIVE DATA VISUALIZATION

AI/MACHINE LEARNING

I like Dutch class!



Galileo

Sorry, Aristotle, you're just wrong.

Aristotle, mon ami, above and below, it's all matter

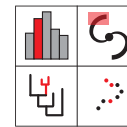
We're gonna need better math.



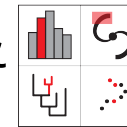
This apple. That moon. Could they feel the same pull?



Seeing the universe, more clearly, with



glue &
multidimensional data exploration
glueviz.org



glupyter
multidimensional data exploration

These slides, along with a link to the Google Drive with all the demos/notebooks, will be posted to my website and to the JupyterCon 2023 Slack.



- 📁 CosmicDS
- 📁 Data+Climate
- 📁 JWST tools
- 📁 Local Bubble
- 📁 Orion
- 📁 Radcliffe Wave (RESTRICTED pending publication embargo)
- 📁 Sibling cluster orbits
- 📁 WWT

MANY thanks to Pat Udomprasert, Jon Carifio, Jonathan Foster, Cami Pacifici, Theo O'Neill, Catherine Zucker, Mike Foley, Ralf Konietzka, Cameren Swiggum, and Peter Williams for providing the ipynb examples!